

The image features a complex abstract graphic design. A white vertical line runs down the center. To the left of this line, there are several overlapping circles and semi-circles in shades of blue, green, and orange. To the right, a large, dark green shape, resembling a stylized eye or a lens, is partially visible. The background is filled with faint, light green text from a Dutch text book, which is mostly illegible but includes words like 'Onderzoek', 'begrip', and 'leerlingen'.

PATRICK ROOIJACKERS

OOG VOOR DIEP BEGRIP

*Onderzoek naar het tekstbegrip
van vwo-leerlingen*

Oog voor diep begrip

*Omdat een ladder niet leunen kan tegen een wolk
is ooit het weerbericht uitgevonden én het boek*

K. Michel

Dit proefschrift werd mede mogelijk gemaakt door financiële steun van het Dudoc-Alfa-programma, een gezamenlijk initiatief van de faculteiten Letteren/Geesteswetenschappen van de Open Universiteit, de Radboud Universiteit Nijmegen, de Rijksuniversiteit Groningen, Tilburg University, de Universiteit Leiden, de Universiteit Utrecht, de Universiteit van Amsterdam en de Vrije Universiteit. Het Dudoc-Alfa-programma biedt eerstegraads bevoegde docenten de mogelijkheid om, naast hun baan in het onderwijs, vier jaar lang een promotieonderzoek uit te voeren op het terrein van de vakdidactiek van de geesteswetenschappen.



Universiteit
Utrecht



ISBN: 978-94-6473-022-7

Copyright: Patrick Rooijackers, Eindhoven
Omslagontwerp & opmaak: Patricia Harsevoort, Amersfoort
Uitgave: Cito, Arnhem
Druk: Ipskamp Printing, Enschede

OOG VOOR DIEP BEGRIP

Onderzoek naar het tekstbegrip van vwo-leerlingen

Focusing on deep reading

*Research on reading comprehension of Dutch pre-university students
(with a summary in English)*

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 17 februari 2023 te 12.15 uur door

Petrus Johannes Henricus Rooijackers
geboren op 14 december 1978
te Deurne

Promotor:

Prof.dr. H.H. van den Bergh

Copromotoren:

Dr. G. van Silfhout

Dr. E. Mantingh

Beoordelingscommissie:

Prof.dr. H. Van Keer

Prof.dr. R.C.M. van Steensel

Prof.dr. P.H.A. Coopmans

Prof.dr. H.C.J. de Graaff

Prof.dr. E.W. van der Knaap

INHOUDSOPGAVE

1 INLEIDING	9
1 De huidige invulling van tekstbegrip historisch ingebed	12
1.1 1900-ca. 1940: 'levend lezen' als uitgangspunt voor de leesles	13
1.2 Ontwikkeling naar tekstbegrip als logische inhoud	15
1.3 Doorwerking van de communicatieve wending in het heden	17
2 Het onderzoek in dit boek: afbakening	19
2.1 Actualiteit, probleemstelling en kader	19
2.2 Doel en onderzoeksvragen	22
2.3 Uitgangspunten bij het onderzoek	24
2.4 Algemene opzet	29
2.5 Enkele opmerkingen over de ontstaansgeschiedenis	29
2 TEKSTEN LEZEN, VRAGEN BEANTWOORDEN, GEOBSERVEERD BIJ VWO 4-LEERLINGEN	31
1 Inleiding	33
1.1 Het tekstbegrip van vwo-leerlingen	33
1.2 Relatie tussen vooraf-lezen en vragen beantwoorden	34
1.3 Onderzoekshypothesen	36
2 Methode	36
2.1 Participanten en experimentopzet	36
2.2 Stimuli	37
2.3 Procedure	38
2.4 Data-analyse	39
3 Resultaten	41
3.1 Vooraf-lezen	41
3.2 Vragen beantwoorden	44
3.3 Relatie tussen vooraf-lezen en antwoord	46
4 Discussie en conclusie	47
4.1 Vooraf-lezen	47
4.2 Vragen beantwoorden	48
4.3 Relatie tussen vooraf-lezen en beantwoorden	49
4.4 Slotobservaties	49

3	TEKSTEN LEZEN, VRAGEN BEANTWOORDEN, GEOBSERVEERD BIJ VWO 2-, VWO 4- EN VWO 6-LEERLINGEN	53
1	Inleiding	55
1.1	Aanleiding	55
1.2	Tekstbegrip van vwo-leerlingen in kaart	56
1.3	Onderzoekshypothesen	59
2	Methode	60
2.1	Participanten en experimentopzet	60
2.2	Stimuli	61
2.3	Procedure	62
2.4	Data-analyse	63
3	Resultaten	66
3.1	Algemeen	66
3.2	Vooraf-lezen	67
3.3	Beantwoorden	71
3.4	Relatie tussen vooraf-lezen en beantwoorden	75
3.5	Relatie tussen vooraf-lezen en het antwoord	76
3.6	Retrospectief interview	76
4	Discussie en conclusie	78
4.1	Samenvatting	78
4.2	Discussie	80
4.3	Kanttekeningen	82
4.4	Implicaties voor het onderwijs	83
4	DE RELATIE TUSSEN HET VOORAF-LEZEN VAN TEKSTEN EN HET BEANTWOORDEN VAN BEGRIPSVRAGEN	85
1	Inleiding	87
1.1	Aanleiding	87
1.2	Tekstbegrip en intentionaliteit	88
1.3	Taakmodel bij een tekst met vragen	90
1.4	Onderzoekshypothesen	92
2	Methode	93
2.1	Experimentopzet en participanten	93
2.2	Voorbereiding op deelname aan experiment	96
2.3	Constructie van het materiaal	96
2.4	Data-analyse	98
3	Resultaten	98
3.1	Algemeen	98
3.2	Controle manipulatie	99
3.3	Antwoordscore	101
3.4	Samenvatting resultaten	103

4	Discussie en conclusie	105
4.1	Kanttelingen	106
4.2	Implicaties voor het onderwijs	107
4.3	Conclusie	110
5	OBSERVEREND LEREN LEZEN OP HET VWO	111
1	Inleiding	113
1.1	Aanleiding	113
1.2	Theorie en onderzoeksvragen	114
2	Methode	117
2.1	Experimentopzet en participanten	117
2.2	Doelen experimentele lessenserie	117
2.3	Constructie materiaal experimentele lessenserie	118
2.4	Voor- en nameting	119
2.5	Data-analyse	120
3	Resultaten	121
3.1	Experimentafname en data-uitval	121
3.2	Voor- en natoetsen	121
3.3	<i>Post-then-pre-test</i>	122
4	Discussie en conclusie	123
6	DE EFFECTIVITEIT VAN EEN LESSENSERIE OVER DIEP LEZEN	127
1	Inleiding	129
1.1	Leesonderwijs in Nederland	129
1.2	Diep lezen centraal	132
1.3	Strategie-instructie	133
1.4	Doelen en onderzoeksvragen	135
2	Methode	137
2.1	Participanten	137
2.2	Experimentopzet	137
2.3	Materiaal	138
2.4	Controle-instrumenten implementatie	143
2.5	Meetinstrumenten effect lessenserie	145
2.6	Data-analyse	147
3	Resultaten	147
3.1	Implementatie van de lessenserie	147
3.2	Evaluatie van de lessenserie	151
3.3	Effect van de lessenserie	152
3.4	Samenvatting resultaten per onderzoeksvraag	154

4	Discussie en conclusie	155
4.1	Kanttekeningen	156
4.2	Wetenschappelijke implicaties	157
4.3	Implicaties voor het onderwijs	158
4.4	Aandachtspunten	159
4.5	Conclusie	160
7	DISCUSSIE	163
1	Doel en opzet van het onderzoek	165
2	De centrale onderzoeksvragen beantwoord	166
2.1	Eerste onderzoeksvraag	167
2.2	Tweede onderzoeksvraag	168
2.3	Derde onderzoeksvraag	171
3	Implicaties voor de wetenschap	172
3.1	Methodologische kanttekeningen	172
3.2	Suggesties voor vervolgonderzoek	178
4	Implicaties voor het onderwijs	185
4.1	Implicatie 1: Maak het lezen van een tekst betekenisvol	186
4.2	Implicatie 2: Reflecteer samen op het begripsproces	190
4.3	Implicatie 3: Vergelijk, weeg en integreer informatie tussen tekst(del)en	193
4.4	Implicatie 4: De rol van de docent in het leerproces is cruciaal	196
4.5	Samenvatting	198
5	Een persoonlijke aanbeveling tot slot	199
	LITERATUUR	203
	BIJLAGEN	223
	SUMMARY	251
	DANKWOORD	263
	NAMENREGISTER	265
	AUTEURSINFORMATIE	269

HOOFDSTUK 1

Inleiding

Huidige leergangen Nederlands voor vwo-leerlingen hanteren in hun leesparagrafen (vooral nog) vaak een opvallend uniforme aanpak. Ze richten zich vaak op ‘strategie-instructie’, waarmee leerlingen wordt geleerd globaal, oriënterend, intensief en kritisch te lezen (bijv. Bolt et al., 2018). ‘Leesvaardigheid’ betekent in leesparagrafen niet zozeer technische leesvaardigheid of vloeiend hardop kunnen lezen, maar vooral ‘tekstbegrip’: het cognitief, rationeel analyseren en begrijpen van de inhoud, vooral op hoofdlijnen (bijv. Frank et al., 2019). Leerlingen moeten vaak één tekst lezen en daarna individueel een set tekstafhankelijke begripsvragen beantwoorden, waarin tekstanalyse een belangrijke rol speelt (bijv. Merckx et al., 2018). Men richt zich daarbij vooral op zakelijke teksten, artikelen uit Nederlandse dagbladen of tijdschriften; het lezen van niet-zakelijke teksten (narratieve teksten, fictie en literatuur) is afwezig (bijv. *Talent*, [2017]) of is ondergebracht in een aparte paragraaf (bijv. Frank et al., 2019). Overigens zijn er zeker enkele afwijkende benaderingen van de les tekstbegrip voorhanden (bijv. Pronk-Van Eunen & De Vos, 2018).

De vraag is in hoeverre deze nogal uniforme aanpak zo vanzelfsprekend is als hij lijkt. Wetenschappelijk gezien is hij dat in elk geval niet. Zo wijkt de vorm van strategie-instructie in de leergangen markant af van de huidige wetenschappelijke standaard (vgl. Almasi & Fullerton, 2012), en dat geldt evenzeer voor de vrij rigoureuze scheiding in tekstbegrip van zakelijke teksten enerzijds en niet-zakelijke teksten anderzijds: die is naar huidige wetenschappelijke inzichten moeilijk te verdedigen (vgl. Kintsch, 1998).

In dit boek staat het tekstbegrip van vwo-leerlingen centraal, de vaardigere lezers in het Nederlandse voortgezet onderwijs (vgl. Expertgroep Doorlopende Leerlijnen Taal, 2008, 2009; Gubbels et al., 2019; Van Silfhout, 2014; etc.). Het richt zich daarbij niet zozeer op lagere-orde-begripsprocessen, zoals het decoderen van woorden, maar op hogere-orde-begripsprocessen, op diep tekstbegrip, oftewel: begrip op het niveau van het situatiemodel, zoals het wordt genoemd in de meest courante leestheorie, de Constructie-Integratie-theorie (Kintsch, 1998): slagen lezers erin op adequate wijze de beweringen in een tekst aan te vullen met én te integreren in hun kennis van de wereld? Dit diepe begrip zullen we beschrijven en analyseren bij leestaken met zakelijke teksten, zoals te vinden in gangbare vwo-leergangen Nederlands. Via dit beschrijvend onderzoek zullen we¹ reflecteren op pijnpunten

¹ Ik zal in dit boek consequent spreken vanuit de we-vorm. Het onderzoek dat is opgenomen in hoofdstuk 2 tot en met 6, is eerder geschreven en gepubliceerd met andere auteurs. Hoewel dit eerste en het slothoofdstuk niet afzonderlijk verschijnen, geldt daarvoor dezelfde werkwijze. Daarom is ‘we’ in alle hoofdstukken op zijn plek.

binnen deze hedendaagse lespraktijk en over verbetermogelijkheden daarin. Op basis daarvan willen we vervolgens een effectieve, alternatieve, meer eigentijdse invulling ontwerpen van de lessen tekstbegrip.

Daarbij staat één belangrijk uitgangspunt voorop: visualisatie van tekstbegrip. Tekstbegrip is een enigszins ongrijpbaar mentaal proces, dat zich voor een lezer (deels) onbewust voltrekt en dat zich voor een toeschouwer alleen indirect laat waarnemen. Daarom zal in ons onderzoek visualisatie van tekstbegrip een belangrijke rol spelen, ten eerste als onderzoeksinstrument – we gebruiken in de hoofdstuk 2 en 3 oogbewegingenonderzoek – en ten tweede als onderwijsinstrument – we gebruiken in hoofdstuk 5 en 6 oogbewegingsfilmpjes om met leerlingen na te denken over het tekstbegrip van leeftijdsgenoten en dat van henzelf. Tekstbegrip van lezers kan daardoor als het ware tastbaar worden, voor onderzoekers, voor lezers zelf én uiteindelijk ook voor leerlingen in klassensituaties.

Voordat we de deelonderzoeken rapporteren in hoofdstuk 2 tot en met 6, zullen we in dit inleidende hoofdstuk inzoomen op een drietal zaken. Ten eerste willen we laten zien dat de huidige sterke gerichtheid op tekstbegrip van zakelijke teksten in de leesles Nederlands historisch gezien zeker niet vanzelfsprekend is: in § 1.1 bespreken we hoe er in de periode 1900-1940 sprake was van een paradigmatisch afwijkende aanpak van de leesles Nederlands, en in § 1.2 gaan we in op de vraag waardoor deze opvatting in de periode 1940-1970 naar de achtergrond verdween. Ten tweede zullen we in § 1.3 tonen hoe de huidige aanpak in leergangen Nederlands vooral gestoeld is op verouderde neerlandistische vakdidactische inzichten uit de jaren zeventig en tachtig van de 20e eeuw. Ten derde zullen we in § 2.1-2.5 uitgebreid stilstaan bij de aanleiding, probleemstelling, doelen, vraagstelling, uitgangspunten, de algemene opzet en de ontstaansgeschiedenis van dit boek.

1 De huidige invulling van tekstbegrip historisch ingebed

We starten eerst met een ideeënhistorische uiteenzetting, waarin we vakdidactische opvattingen beschrijven over tekstbegripsonderwijs in het schoolvak Nederlands in het Nederlandse voortgezet (middelbaar) onderwijs, gedurende de periode van 1900 tot nu: we zullen een drastisch andere opvatting over tekstbegrip schetsen dan vandaag de dag doorgaans wordt onderschreven. In deze uiteenzetting gaan we vrijwel alleen in op uitlatingen van vakdidactici en publicerende docenten Nederlands, waarbij we ons in hoofdlijnen zullen baseren op de studies van Hulshof et al. (2015), Van Kalmthout (2002) en meer zijdelings Van de Ven (1996).

De geschiedenis van het Nederlandse tekstbegripsonderwijs *an sich* zal hiermee dus niet worden getypeerd. We schetsen niet zozeer het uitgevoerde curriculum rond tekstbegrip, ‘de werkelijkheid’ van de gemiddelde leesles van ca. 1920 zogenoemd, maar het curriculum rond tekstbegrip zoals docenten en vakdidactici Nederlands het zelf aangeven in de leesles na te streven. Deze schets kan ons helpen scherper uitgangspunten, kaders en beperkingen van huidige opvattingen over tekstbegrip en tekstbegripsonderwijs te zien; ze geeft reliëf, zogenoemd, aan de ver-

onderstellingen en uitgangspunten in het leesonderwijs van vandaag de dag, zodat we deze als minder vanzelfsprekend kunnen waarnemen. Overigens: wat er daadwerkelijk in deze periode in het klaslokaal Nederlands gebeurde, is een interessante, complexe vraag die een andere, veel systematischere studie vergt dan we hier zullen en kunnen bieden: er is namelijk geen systematisch observatieonderzoek uit de periode 1900-1940 beschikbaar. Aan de beantwoording van die enkele vraag zou een heel proefschrift kunnen worden gewijd. Hier volstaan we daarom met een schets van wat men zegt na te willen streven in het leesonderwijs.

1.1 1900-ca. 1940: 'levend lezen' als uitgangspunt voor de leesles

In de handboeken en leergangen voor het schoolvak Nederlands uit de periode 1900-1940 wordt leesvaardigheid bepaald anders benaderd dan in huidige leergangen Nederlands. Vooral literatuur staat erin centraal. Literatuur werd vanuit het negentiende-eeuwse ideaal van *Bildung* van groot nut geacht voor de taalkundige, nationalistische, esthetische en morele vorming van leerlingen; aan zakelijke (journalistieke en wetenschappelijke) teksten werd in de leesles geen wezenlijke rol toegekend (Hulshof et al., 2015; Van Kalmthout, 2002). Letterkundige bloemlezingen bepaalden in de periode 1900-1940 dan ook het leesonderwijs.

In deze periode domineert in het Nederlandse moedertaalonderwijs een "individueel-expressieve" (Hulshof et al., 2015) of "individueel-esthetische opvatting" (Van de Ven, 1996), waarin de persoonlijke beleving van literatuur door leerlingen centraal stond. Deze visie op leesonderwijs werd rond 1900 geïntroduceerd door de Zeeuwse HBS-docent J.H. van den Bosch, die het spreken in het onderwijs alomtegenwoordig maakte vanuit het motto 'Taal is klank' (bijv. Van den Bosch, 1909). We typeren deze visie hier kort in vier kenmerken.

Ten eerste: het lezen van een tekst was volgens prominente vakdidactici uit de periode 1900-1940 vooral een activiteit van de ziel, een associatief, creatief, emotioneel proces. Taal, zo verkondigde bijvoorbeeld de taalkundige Van Ginneken in 1917, "is de gewilde verbinding van menselijk geluid met menselijken zielsinhoud", en door verklanking wordt een tekst in de ziel van een toehoorder "warm [...] van tintelend bewustzijn" (geciteerd in Van Kalmthout, 2002, p. 110-111). Docenten dienden leerlingen daarom deze zielssensatie in de leesles te laten ervaren. Volgens de Bredase leraar Nederlands en letterkundige J. Mathijs Acket moest een docent in de leesles de "associatie" bij leerlingen bevorderen: "Op de eerste plaats is nodig dat [de docent] zelf heel erg meeleeft, met ziel en geest, zodat-ie geboeid wordt en verrassingen ondergaat. En die boeiing spreekt-ie uit met korte woordjes, haast zonder 't lezen te onderbreken." Het gevolg is dat de belangstelling van leerlingen wakker wordt: "dat suffice passieve lezen wordt 'n daad van de Wil, ze willen hun portie hebben van 't geen er te halen is uit dat boek" (Acket, 1912, p. 73).

Ten tweede: leesvaardigheid wordt in deze opvatting gedefinieerd als het beziend en welluidend voordragen van een tekst; leerlingen moesten vooral leren "levend te lezen" (Van Kalmthout, 2002). Leerlingen dienden een literaire tekst vloeiend en gloedvol te kunnen reciteren. Dan lieten ze namelijk zien dat ze de tekst begrepen,

was de gedachtegang. Wanneer ze “de betekenis-waarde van intonatie en accent, van tempo en rust, van gebaar en mimiek” tijdens lezen hebben kunnen vatten, dan “spreekt het vanzelf dat [leerlingen] daartoe de tekst begrijpen moeten”, zo stelt bijvoorbeeld de Nijmeegse leraar Nederlands en vakdidacticus J.G.M. Moormann (1936, p. 58). Hardop-lezen is daarbij essentieel; aan individueel ‘stil-lezen’ kent men een geringe waarde toe (vgl. Moormann, 1936; Kramer, 1938).

Ten derde: analytisch, rationeel, strikt op inhoud en cognitie gericht lezen had volgens deze vakdidactici en docenten weinig in de leesles te zoeken; dit stond het ‘levend lezen’ al snel in de weg. Wanneer taal teruggebracht werd tot louter denken, dan raakte de samenhang tussen taal en klank al snel in de knel. De Utrechtse vakdidacticus W. Kramer stelde bijvoorbeeld dat wie zich “uitsluitend richt op het geschreven woord”, het “skelet voor het lichaam houdt”. En zo’n lezer belandt al snel op “de alleenzigmakende weg tot het abstracte denken, waar alle natuurlijk taal-leven sterft” (Kramer, 1938, p. 276).

Ten vierde: de persoonlijke interpretatie van lezers stond in deze benadering voorop. De rationele tekstverklaring zou volgens docenten en vakdidactici uitgaan van een eenduidige interpretatie van teksten, een interpretatie die bovendien hiërarchisch door de docent aan leerlingen werd opgelegd – en dat was geen goede zaak. Zo meent bijvoorbeeld de Arnhemse leraar Nederlands F.G.A. Stemvers:

Alle begrijpen is subjectief, wisselend van persoon tot persoon, van levensjaar tot levensjaar. Wanneer men dan beweert dat een tekst ‘volkomen door allen begrepen wordt’ bedoelt men: dat alles door allen begrepen is, *zoals de docent zelf meent dat het begrepen moet worden* [curs. Stemvers]. Maar wie zegt hem, dat hijzelf de tekst juist begrepen heeft, vooral als dit een niet al te nuchtere tekst is. Wie zegt hem ook, of er met dit sterk analyserend begrijpen niet iets anders verloren is gegaan? (Stemvers, 1938, p. 102)

Met deze observaties sluit Stemvers opvallend aan bij contemporaine literatuurwetenschappelijke opvattingen in het buitenland, vooral Rosenblatt (1938). De ideeën van de laatste zullen echter pas vanaf de jaren zeventig luide weerklank vinden, zoals in *reader response criticism* (bijv. Holland, 1975), en ze spelen nog altijd een belangrijke rol in huidige literatuurwetenschappelijke inzichten (vgl. Brillenburg Wurth & Rigney, 2019).

Overigens is het de vraag in hoeverre de meeste docenten Nederlands uit het middelbare onderwijs in deze periode deze romantisch-esthetische opvattingen aanhingen en praktiseerden; volgens Hulshof et al. (2015) betrof het hier vooral een luidruchtige minderheid van publicisten rondom het schoolvak. In de praktijk lijken docenten vaak klassikaal een literaire tekst te lezen, terwijl ze leerlingen lukraak bevroegen op woordbetekenissen, grammaticale, taal- en letterkundige feiten; zo werd tekstbegrip ook getoetst (vgl. Van Kalmthout, 2002, p. 107). In handboeken en publicaties maakte de romantisch-esthetische opvatting echter nog lang school: het cruciale belang van het beziel, hardop lezen is nog lang na de jaren dertig terug te vinden in vakdidactische publicaties. De tot in 1971 (!) herdrukte, veelgebruikte *Didactische handleiding voor de docent in de moedertaal* (1962) van

Van Dis en anderen noemt goed hardop lezen nog altijd van “essentiële betekenis” in het middelbaar onderwijs omdat het “directer dan het stil-lezen” de stijl ten goede kan beïnvloeden en als “wél lezen” de smaak kan helpen bevorderen (Van Dis et al., 1962, p. 88; vgl. Van den Ent, 1941).

In dit tijdvak is men in publicaties dus wars van een strikt cognitieve gerichtheid op de inhoud van teksten en stelt men juist de (esthetische) persoonlijke beleving van een tekst voorop; men beschouwt hardop-lezen als essentieel en keert zich af van de eenduidige interpretatie van teksten – uitgangspunten die tegengesteld zijn aan de uitgangspunten in leesparagrafen in huidige leergangen Nederlands voor het voortgezet onderwijs.

1.2 *Ontwikkeling naar tekstbegrip als logische inhoud*

We signaleerden het al: de tijden zijn veranderd. Literatuur neemt vandaag de dag in de les Nederlands weliswaar nog altijd een prominente plek in, maar de aandacht voor zakelijke teksten is de afgelopen twintig jaar sterk toegenomen (Witte, 2018). Hardop klassikaal lezen heeft in de bovenbouw sinds de jaren negentig nauwelijks nog een plaats binnen het bovenbouwcurriculum van Nederlandse havo- en vwo-scholieren; stil-lezen is de standaard (vgl. CVEN, 1991; Meestringa & Ravesloot, 2012) – al zijn er plannen hardop-lezen weer een plek in het curriculum Nederlands te geven (SLO, 2021b). In deze paragraaf staan we kort stil bij de vraag hoe deze gewijzigde visie op leesvaardigheid tot stand is gekomen.

Vanaf de jaren dertig begonnen er vanuit vakdidactici krachtige tegengeluiden te klinken tegen de romantisch-esthetische visie op leesonderwijs. Allereerst kwam er vanaf 1930 in het onderwijs meer oog voor de maatschappelijke relevantie van niet-literaire teksten. Een prominent tegenstander van het romantisch-esthetische paradigma was J. Leest, directeur van de Rijks-HBS te Gouda. Hij beschouwde het in zijn vakdidactisch handboek *Het voortgezet onderwijs in de moedertaal* (1932) als “een der hoofdfouten van het voortgezet moedertaalonderwijs in ons land” dat het “aan talrijke scholen bijna uitsluitend een voorbereiding wil zijn tot het genieten van literaire kunst” (Leest, 1932, p. 21). Leest wees op het grote maatschappelijke belang van vaardigheden waarmee leerlingen journalistieke, wetenschappelijke, ambtelijke en handelstaal konden doorgronden. Er komen in deze periode dan ook leergangen beschikbaar voor de les Nederlands die ruim plaats bieden aan wetenschappelijk en journalistiek proza, zoals *Verstandelijke taal* van J. Leest zelf (1932) en het veelgebruikte *Nederlands na 1600* van Karssemeijer en Kazemier (1938) (Hulshof et al., 2015).

Bovendien kwam er vanaf de jaren dertig in het onderwijs steeds meer oog voor de (meta)cognitieve ontwikkeling van kinderen (Stellwag, 1955). Een prominente stem hierin was de vakdidacticus en pedagoog Langeveld. Deze brak in zijn dissertatie *Taal en denken* (1934) een lans voor meer analyse en logisch redeneren in de leesles: “de grondvorm van de taalbeheersing [is] toch zeer eenvoudig. Wat kan zij anders zijn dan het pogen verstaan, *grondig* verstaan van een logies streng opgebouwde samenhang in goed Nederlands geschreven?” De toenmalige leesles kon

dit volgens hem niet geven: daarin werden met “allerlei verhaaltjes” niet veel meer “dan wat vage, zwakke of eenzijdige emotietjes gewekt” (Langeveld, 1934, p. 146). Langeveld beschouwde de individuele ‘stilleesoefening’ met begripsvragen als een essentiële toevoeging aan de leesles, omdat ze dwingt tot “exakt taalverstaan en zich richt op ‘logiese samenhang’” en ze daarmee ontbrekend begrip bij leerlingen kan blootleggen (Langeveld, 1934, p. 188). Het individueel, in stilte tot je nemen van een tekst, vooral gericht op de inhoud ervan, maakt in de jaren dertig dan ook een opmars in het onderwijs. De “triumf van het stil-lezen”, noemt Diels (1938) het zelfs. Als toets krijgt de stilleesoefening vanaf eind jaren dertig een centrale plek in het lager onderwijs (Haas, 1995) en vanaf 1945 in het middelbaar onderwijs (Riemens-Feteris, 1971).

Vooraf ook de zogenaamde ‘communicatieve’ of ‘normaal-functionele’ wending in het schoolvak Nederlands, die inzet vanaf de jaren zeventig, heeft de drastische wijziging in opvattingen rond leesvaardigheid beïnvloed (zie met name Ten Brinke, 1976; vgl. Van de Ven, 1996).² Volgens deze markante heroriëntatie op het schoolvak Nederlands diende het moedertaalonderwijs zich vooral te richten op algemene taalvaardigheden. Vaardigheden waren immers cruciaal in een maatschappij waarin niet meer de afkomst maar de individuele verdienste telde (Hulshof et al., 2015). Ze vertaalde zich vanaf 1970 in drastisch andere accenten in het klaslokaal: het taalgebruik van leerlingen zelf kwam centraal te staan, literatuuronderwijs moest meer emancipatorische doelen gaan dienen, en zakelijke teksten uit het alledaagse leven kregen een vooraanstaande plek (Ten Brinke, 1976). Bekend is bijvoorbeeld de uitgebreide leesmethode *Taaldaden* (1979) van de Leidse taalbeheerser en voormalig docent Nederlands Braet waarin leerlingen de betekenis, opbouw, bedoeling en publieksgerichtheid van zakelijke teksten analyseren zodat ze “moeilijke teksten van beroepsschrijvers” leren te begrijpen en “kritisch” te benaderen (Braet, 1979, p. 5).

In de jaren zeventig raakt het lezen van zakelijk teksten in het schoolvak Nederlands gescheiden van het lezen van literatuur: de taalbeheersersdiscipline en de literatuurwetenschappelijke discipline in de neerlandistiek gaan hun eigen koers varen (Hulshof et al., 2015). Veelzeggend is het vakdidactische handboek *Zeggen-schap* van Griffioen (1975) waarin het lezen van zakelijke en van niet-zakelijke (literaire) teksten in twee aparte hoofdstukken worden ondergebracht. Deze keuze wordt in dit handboek expliciet onderbouwd met inzichten uit *Literatuurwetenschap* (1970) van de Utrechtse literatuurwetenschapper F.C. Maatje: zakelijke teksten zijn namelijk ‘referentieel’ – bij deze teksten is het relevant dat de lezer de werkelijkheid buiten de tekst controleert – en staan dus een eenduidige bespreking toe in de leesles; literaire teksten zijn daarentegen ‘niet-referentieel’ – controle van de werkelijkheid buiten de tekst is irrelevant – en vragen dus in de leesles om interpretatie (Griffioen, 1975, p. 279-280). Daarom zouden ze ook een afwijkende didactiek van docenten vragen en is het logisch de behandeling van deze twee tekstsoorten in aparte hoofdstukken onder te brengen.

² Hier schetsen we de ontwikkeling binnen de neerlandistiek. Echter, ook internationaal is er in de onderwijswetenschappen een verschuiving naar het communicatieve domein te zien (zie Slagter, 2001).

Deze scherpe scheiding tussen beide tekstsoorten wordt daarna snel gangbaar in het schoolvak Nederlands, zeker wanneer het ‘communicatief-emancipatoire paradigma’ in de jaren tachtig verder verschuift naar het ‘communicatief-utilitaire paradigma’ (Van de Ven, 1996): het belang van functionele, alledaagse taalvaardigheid van leerlingen wordt steeds sterker benadrukt, waarbij het emancipatoire verheffingsideaal naar de achtergrond verdwijnt. In de jaren negentig zal de scheiding tussen zakelijk lezen (taalbeheersers) en literatuur (letterkundigen en literatuurwetenschappers) onuitgesproken en vanzelfsprekend het vertrekpunt vormen voor toekomstige vernieuwingen rondom het schoolvak Nederlands (CVEN, 1991).

1.3 Doorwerking van de communicatieve wending in het heden

Vooropgesteld: er zijn in de periode 1985-2020 zeker wijzigingen opgetreden in het domein zakelijk lezen, zoals de opname van argumentatieve vaardigheden vanaf 1998 in de eindtermen voor leesvaardigheid. (Zie voor de achtergronden bij de procedureel opmerkelijke invoering van dit subdomein: Van der Aalsvoort, 2016.) Echter, wie de leesparagrafen in de meest gebruikte leergangen Nederlands van nu doorneemt, ziet nog altijd veel kenmerken die zijn terug te voeren op ideeën over tekstbegrip uit de periode 1970-1985. We zullen er een aantal malen in dit boek nog over te komen spreken. We noemen er hier drie: de nogal eigenzinnige vorm van ‘strategie-instructie’, de tekst-met-vragen-aanpak en de vrij strikte scheiding tussen zakelijke en niet-zakelijke teksten.

Ten eerste: de invulling van strategieonderwijs in veelgebruikte hedendaagse leergangen Nederlands als *Nieuw Nederlands* en *Op niveau* en in een vakdidactisch handboek als Bonset et al. (2019) wijkt nogal af van de internationale standaard. Onder het etiket ‘strategieën’ wordt in leergangen doorgaans een set generieke manieren van lezen aangeboden: zoekend, globaal, oriënterend, precies en studierend lezen. Internationaal wordt een strategie echter niet beschouwd als zo’n generieke leesaanpak, maar als een bewust, doelgericht middel dat lezers kunnen inzetten om hun begrip aan te passen en te verbeteren (Afflerbach et al., 2008), zoals ‘herlezen van kernzinnen’, ‘bepalen van de hoofdgedachte’, ‘activeren van voorkennis’, etcetera (bijv. Almasi & Fullerton, 2012; Pressley & Afflerbach, 1995).

Waar komt dit afwijkende strategie-concept van leergangen Nederlands dan vandaan? Ze blijkt terug te voeren op de ‘SQ3R’-studeermethode uit het boek *Effective study* (1946) van de Amerikaanse onderwijspsycholoog F.P. Robinson: met deze methode leerden Amerikaanse universitair studenten stapsgewijs een studietekst te bestuderen (SQ3R: *Survey, Question, Read, Recite and Review*). In Nederland belandde deze studeermethode vervolgens in een handboek voor studenten, *Effectief studeren* van Van Parreren et al. (1969). Daarna werd ze nogal eigenzinnig herzien door de taalbeheersers Drop en De Vries in *Taalbeheersing* (1974); deze zetten haar om in een set inhoudelijk losstaande, apart onderwijsbare ‘leesstrategieën’ (zoekend lezen, oriënterend lezen, etcetera). Deze bewerking kwam vervolgens terecht in didactische handboeken voor het schoolvak Nederlands – zij het met kanttekeningen: “Zo zullen er nog wel meer strategieën zijn en waarschijnlijk

heeft ieder ook z'n eigen strategieën" (Griffioen, 1975, p. 291). Vrijwel ongewijzigd vigeerde deze indeling daarna tot 2006 (sic!) in de kerndoelen voor het voortgezet onderwijs, en ze is tot op de dag van vandaag in de leesparagrafen van veel leergangen Nederlands te vinden (zie Rooijackers et al., 2021b).

Het tweede voorbeeld is de prominente plek van de oude stilleesoefening, alias de 'tekst met vragen', in huidige leergangen Nederlands havo/vwo. De dominantie van deze taak in het tekstbegripsonderwijs Nederlands wordt inmiddels al decenia bekritiseerd (bijv. Hoogeveen & Bonset, 1998; Westhoff, 2012). Wie de leesparagrafen van hedendaagse veelgebruikte leergangen Nederlands voor het voortgezet onderwijs doorneemt zoals *Nieuw Nederlands* en *Op niveau*, ziet een vrijwel identieke inrichting van deze leestaak: een voorkeur voor een top-downbenadering; één tekst uit een gezaghebbend dagblad of tijdschrift staat centraal; een realistische context of betekenisvol doel ontbreekt bij de opdracht; tekst(on)afhankelijke vragen en studerend lezen spelen een hoofdrol; leerlingen werken individueel (zie Rooijackers, 2021a). Deze benadering wijkt nogal af van bijvoorbeeld moderne Angelsaksische leesdidactiek (bijv. Fisher & Frey, 2015): daarin worden leraren getraind vragenderwijs klassengesprekken te voeren over de inhoud van een tekst, om zoveel mogelijk diepgang en kwaliteit in de discussie over een tekst te brengen (vgl. Wilkerson & Son, 2011). Een klassieke valkuil bij het stellen van begripvragen – louter controleren of leerlingen een goed of fout antwoord geven (vgl. Durkin, 1978) – ligt bij deze neerlandistische benadering snel op de loer.

Wie op zoek gaat naar de theoretische fundering van deze benadering, komt al snel uit bij *Leesvaardigheid Nederlands: Omgaan met zakelijke teksten* (1994) van de taalkundigen Hendrix en Hulshof. In deze studie, lange tijd de enige zelfstandige vakdidactische publicatie over het tekstbegripsonderwijs in het schoolvak Nederlands, wordt een didactiek geschetst waarbij leerlingen via intensieve training met tekst(on)afhankelijke bevragingstechnieken zelfstandig en 'diep' leren lezen. Hun "interactieve analyse" kenmerkt zich door een gerichtheid op het begrijpen van één enkele gezaghebbende tekst, een zekere voorkeur voor een top-downanalyse en een vaststaand generiek leesdoel: studerend lezen. Via een vijf-stappenprocedure, waarin 'aandachtsturende richtvragen' een belangrijke rol spelen, leren leerlingen de inhoud van een tekst bloot te leggen. Hendrix en Hulshof baseren zich voor deze interactieve benadering grotendeels op de opvattingen ("zeer behartenswaardig") uit het handboek *Taalbeheersing* van Drop en De Vries (1974) en uit *Instrumentele tekstanalyse* van Drop (1979); daarnaast plaatsen ze *expressis verbis* hun vraagegestuurde benadering in een neerlandistische traditie uit de jaren zeventig en tachtig (Hendrix & Hulshof, 1994, p. 19; p. 31).

Moderne leergangen Nederlands laten zich – jammer genoeg – nauwelijks tot niet uit over de theoretische fundamenten achter hun didactiek, maar de leesparagrafen in een prominente leergang als *Nieuw Nederlands* bleken theoretisch in 2020 vooral op Hendrix en Hulshof (1994) gebaseerd.³ Ook een handboek als

³ Dat gaf de eindredacteur van de bovenbouweditie van *Nieuw Nederlands* in een persoonlijke mail (27 januari 2020) aan.

Nederlands in de onderbouw (Bonset et al., 2019) kan hierin een rol hebben gespeeld: weliswaar keert dit handboek zich expliciet af van een te sterke nadruk op de tekst-met-vragen-aanpak, maar het grijpt ook met regelmaat terug op deze aanpak bij de beschrijving van mogelijke lees oefeningen en -toetsen.

Het derde voorbeeld is de rigoureuze scheiding tussen het lezen van enerzijds literaire, fictionele en narratieve teksten en anderzijds zakelijke teksten in het huidige schoolvak Nederlands. In het schoolvak Nederlands is deze scheiding tegenwoordig zeer courant: ze is bijvoorbeeld verankerd in twee gescheiden subdomeinen in het referentiekader Taal, de beschrijving van de beheersingsniveaus voor Nederlandse leerlingen (Expertgroep Doorlopende Leerlijnen Taal, 2008, 2009), en ze is ook verankerd in de eindtermen voor het voortgezet onderwijs (CvTE, 2022a, 2022b). Internationaal gezien is ze minder vanzelfsprekend. In het Angelsaksische leesonderwijs worden zakelijke en fictionele/narratieve teksten nauwelijks of niet van elkaar onderscheiden (Fisher & Frey, 2015), en evenmin in de meest gangbare theorie over tekstbegrip, de Constructie-Integratie-Theorie (Kintsch, 1998): de begripsprocessen bij narratieve teksten onderscheiden zich hierin slechts in geringe mate van de processen bij zakelijke teksten. En deze scheiding is ook niet zonder problemen. Zo zouden Nederlandse leerlingen volgens het Referentiekader in fictionele teksten wél stijlmiddelen kunnen herkennen maar niet in zakelijke teksten, op ironie na – terwijl deze stijlmiddelen een prominente rol kunnen spelen in zakelijke teksten (bijv. Braet, 2007).

Kortom, veel kenmerkende opvattingen over tekstbegrip uit de jaren zeventig en tachtig werken nog altijd door in de leeslessen Nederlands in het voortgezet onderwijs van vandaag de dag. We zullen er in dit boek nog een aantal malen op terugkomen: veel aandachtsgebieden uit het internationale leesonderzoek van de laatste twintig jaar zijn nog in beperkte mate doorgedrongen tot de vakinhoud en -didactiek in het leesonderwijs binnen het schoolvak Nederlands, zoals de functionele inbedding van tekstbegrip (McCrudden & Schraw, 2007) en inzichten rondom meervoudig-tekstbegrip (*multiple text comprehension*), waaronder het belang van betrouwbaarheid en bruikbaarheid van informatie (Britt & Rouet, 2012).

2 Het onderzoek in dit boek: afbakening

2.1 Actualiteit, probleemstelling en kader

Het goed kunnen begrijpen van zakelijke teksten heeft de laatste decennia waarschijnlijk alleen maar aan belang gewonnen. Het leren van informatieve teksten is een van de meest basale vaardigheden in de huidige informatiesamenleving (Cattrysse et al., 2017; OECD, 2019), en door de enorme toevoer van informatie via internet en sociale media, heeft de moderne jongere tegenwoordig te lijden onder een ‘infodemie’ (WHO, 2022). Het tekstbegrip van Nederlandse leerlingen staat daarbij sinds 2019 sterk in de media in de belangstelling (bijv. Copier, 2020; Kalshoven, 2019; Truijens, 2020; Welgraven & De Vos, 2019) en Nederlandse overheids-

instanties uitensindsdien zorgen over het tekstbegrip van Nederlandse leerlingen (bijv. Onderwijsraad & Raad voor Cultuur, 2019; Pereira & Nicolaas, 2019; Inspectie van het Onderwijs, 2022). Vooral in de media is de rode vlag sinds 2019 regelmatig gehesen: er zou zich “een leesramp” voltrekken (Truijens, 2020), en het hele Nederlandse stelsel van leesonderwijs en leesbevordering zou “rot” zijn (Van Dijk & Klaver, 2022).

Belangrijke aanjager hierbij is het dalende tekstbegrip van Nederlandse leerlingen op met name de internationaal vergelijkende toets PISA: waar Nederlandse vijftienjarige leerlingen lange tijd gemiddeld tot de betere lezers in Europa behoorden, zakken ze blijkens deze test inmiddels terug naar de middenmoot; vooral op het subdomein evalueren en reflecteren blijken ze minder sterk dan Europese leeftijdsgenoten (OECD, 2019). Dit laatste resultaat wordt weerspiegeld in de bekende PIRLS-test naar het begrip van basisschoolleerlingen: Nederlandse groep 6-leerlingen zijn aanmerkelijk sterker in het vinden van informatie dan in het reflecteren op de tekstinhoud (Gubbels et al., 2017).

In hoeverre de leesvaardigheid van Nederlandse leerlingen daadwerkelijk is gedaald, is een moeilijke vraag. Een recente peiling van de Onderwijsinspectie liet voor 2021 weliswaar een lichte daling in leesvaardigheid van basisschoolleerlingen zien ten opzichte van een peiling van tien jaar eerder, maar de inspectie tekende hierbij wel aan dat lesuitval en schoolsluitingen door de Covid19-epidemie in 2020 en 2021 in deze resultaten waarschijnlijk een belangrijke rol spelen (Inspectie van het Onderwijs, 2022). De PIRLS-test laat vooralsnog geen duidelijke daling in leesvaardigheid zien (Gubbels et al., 2017). En bij de ogenschijnlijk alarmerende resultaten van de PISA-test dient een belangrijke kanttekening geplaatst: deze test is sinds 2010 het subdomein evalueren en reflecteren afwijkend gaan invullen, met meer oog voor meervoudig-tekstbegrip (*multiple text comprehension*), terwijl het Nederlandse onderwijs van oudsher vooral inzet op enkelvoudig-tekstbegrip (*single text comprehension*) (Bonset, 2020; Gubbels, 2020; Van den Broek et al., 2021). Het probleem van de dalende leesvaardigheid van Nederlandse leerlingen ligt daardoor waarschijnlijk al met al genuanceerd. Van den Broek et al. (2021) concluderen in een notitie aan de Tweede Kamer hierover bondig:

Het lijkt erop dat het Nederlandse onderwijs in vergelijking met andere landen relatief goed functioneert wat basisbegripsvaardigheden betreft, maar dat, nu de [PISA-]test de lat wat integreren/evalueren/reflecteren betreft hoger legt, de Nederlandse leerlingen de daarvoor benodigde vaardigheden niet adequaat lijken te beheersen. (Van den Broek et al., 2021, p. 13)

De noodkreten in de media (‘leesramp’) worden door onderzoeksgegevens dus vooralsnog niet krachtig ondersteund. Vooral voor het subdomein evalueren en reflecteren lijkt er werk aan de winkel binnen het schoolvak Nederlands: er lijkt een duidelijke noodzaak dit subdomein beter te verankeren in het leesonderwijs dan nu het geval is.

Daarnaast - en deels hiermee samenhangend - loopt er in het Nederlandse onderwijsveld momenteel een discussie of het leesonderwijs in het Nederlands op

de middelbare scholen fundamenteel moet worden herzien (bijv. Inspectie van het Onderwijs, 2019; Onderwijsraad & Raad voor Cultuur, 2019; Pereira & Nicolaas, 2019). Vanuit de wetenschap zijn er recent zorgen geuit over het Nederlandse leesonderwijs en voorstellen gedaan tot verbetering (Van den Broek et al., 2021; Houtveen et al., 2019; vgl. voor Vlaanderen: Van den Branden et al., 2019). Op het moment van het verschijnen van dit boek loopt er een herziening van het examenprogramma voor het schoolvak Nederlands, die andere accenten en aanvullingen zal opleveren in de landelijke beschrijvingen van de beheersingsniveaus en examenprogramma's voor leesvaardigheid (vgl. Curriculum.nu, 2019; Rijksoverheid, 2022). In dit boek zal deze context nadrukkelijk worden verdisconteerd: we zullen in de discussieparagrafen in hoofdstuk 2 tot en met 6 uitgebreid stilstaan bij de implicaties van deze studies voor met name de bovenbouw (havo en) vwo; ook in het slothoofdstuk staan we uitgebreid stil bij de vraag naar de implicaties van ons onderzoek voor de leesles Nederlands in de bovenbouw.

Verder is er de afgelopen twintig jaar weliswaar relatief veel onderzoek verricht naar het tekstbegrip van Nederlandse scholieren, maar er zijn nog heel wat lacunes in onze kennis. Zo is het tekstbegrip van vmbo-scholieren weliswaar relatief scherp in kaart gebracht (bijv. De Milliano, 2013; Land, 2009; Schram, 2007; Van Silfhout, 2014), maar longitudinaal onderzoek naar hun ontwikkeling in tekstbegrip ontbreekt goeddeels. Er is ook internationaal leesonderzoek zoals PISA (Gubbels et al., 2019), dat 15-jarige vmbo- en havo- én vwo-leerlingen qua tekstbegrip internationaal op rang ordent. Maar er zijn slechts enkele studies die het tekstbegrip van (onder meer) vwo'ers onderzoeken (bijv. Van Silfhout et al., 2015). En ook voor vwo'ers ontbreken longitudinale of zelfs cross-sectionele studies naar hun ontwikkeling in tekstbegrip. Er lijkt dus zeker reden om specifiek *déze* groep leerlingen te bestuderen.

Bovendien is er momenteel een duidelijke roep om meer aandacht voor 'diep lezen' in het leesonderwijs. Jongeren in Nederland zouden weliswaar "volop korte tekstjes [lezen] – berichten op hun smartphone of samenvattende stukjes in schoolboeken – maar [ze] besteden minder tijd aan 'diep lezen': het geconcentreerd lezen van langere teksten of boeken. Mede hierdoor gaat hun leesvaardigheid achteruit" (Onderwijsraad & Raad voor Cultuur, 2019, p. 2; vgl. Ministerie van Onderwijs, Cultuur en Wetenschappen, 2020; Pereira & Nicolaas, 2019). In het basisonderwijs is er inmiddels duidelijk aandacht voor dit diepe begrip in de leesles (vgl. Lapp et al., 2018; Van de Mortel & Ballering, 2022). Er lijkt daarom voldoende reden om ook voor het voortgezet onderwijs een alternatieve effectieve leesaanpak te ontwikkelen die zich vooral richt op dit diepe lezen (vgl. voor het literatuuronderwijs: Dera & Van Steensel, 2022).

In ons onderzoek richten we ons op het begrip van zakelijke teksten door Nederlandse vwo-leerlingen. Het poogt met 21e-eeuwse onderzoekstechnieken zoals eye-tracking-methodologie (Holmqvist & Andersson, 2017) cross-sectioneel een beschrijving te geven van het begrip van vwo-leerlingen. Daarmee hopen we ook aandacht te kunnen schenken aan mogelijke tekortkomingen in de huidige vakdidactiek en -inhoud. Het beschrijft en beproeft bovendien alternatieve benaderingen van tekstbegrip, gebaseerd op recente internationale wetenschappelijke

inzichten (bijv. Britt et al., 2018; McCrudden et al., 2011). Het onderzoek is daarbij steeds sterk vakinhoudelijk en -didactisch van aard: al zullen her en der in dit boek meer theoretische onderzoeksdimensies worden aangestipt, de relevantie ervan voor het schoolvak Nederlands in het voortgezet onderwijs staat voorop, en dan met name voor de vwo bovenbouw.

Het onderzoek in dit boek komt voort uit het Dudoc-Alfa-programma. Dit programma biedt docenten uit het voortgezet onderwijs de mogelijkheid om, naast hun baan in het onderwijs, vier jaar lang een promotieonderzoek uit te voeren op het terrein van de vakdidactiek van de geesteswetenschappen. Het programma streeft daarbij naar valorisatie: het wil een brug slaan tussen de universiteit en het voortgezet onderwijs en het streeft naar een kwaliteitsverbetering en vernieuwing van de alfavakken (Vakdidactiek Geesteswetenschappen, 2022). Basis voor het onderzoek in dit boek vormde een projectvoorstel dat eind 2016 werd ingediend. In september 2017 kon met het onderzoek worden aangevangen.

Met het projectvoorstel uit 2016-2017 lagen de kaders voor dit boek van meet af aan vast. Allereerst: de doelgroep. Het onderzoek zou zich vooral richten op het tekstbegrip van vwo-bovenbouwleerlingen, de vaardigere lezers in het voortgezet onderwijs (vgl. Expertgroep Doorlopende Leerlijnen Taal, 2008, 2009; Gubbels et al., 2019; Van Silfhout, 2014; etc.). Ten tweede: het aandachtsgebied. Dat zou zich richten op diep begrip, begrip op het niveau van het situatiemodel, zoals de Constructie-Integratie-theorie het noemt (Kintsch, 1998). Ten derde: we zouden dit onderzoeken bij zakelijke teksten, aansluitend bij de staande praktijk in de lessen tekstbegrip in het schoolvak Nederlands. Ten vierde: dit alles zouden we onderzoeken binnen de context van het schoolvak Nederlands. Die kaders staan nog altijd, met alle beperkingen van dien. We richten ons dus niet op bijvoorbeeld vmbo-leerlingen en we laten narratieve en fictionele teksten buiten beschouwing. Deze kaders hebben echter als belangrijk voordeel dat we ons ermee kunnen concentreren op de vraag in hoeverre (vaardige) lezers bij de huidige leestaken in het schoolvak Nederlands diep begrip opbouwen.

2.2 Doel en onderzoeksvragen

Het doel van het onderzoek in dit boek is tweeledig. Enerzijds willen we in dit boek lees- en antwoordprocessen beschrijven van huidige vwo-(bovenbouw)leerlingen bij zakelijke teksten, vanuit de bestaande praktijk in het tekstbegripsonderwijs in het schoolvak Nederlands, én we willen deze praktijk tegen de loep houden: in hoeverre bouwen leerlingen bij de huidige leesopdrachten (diep) begrip op van een zakelijke tekst? Anderzijds willen we een alternatieve effectieve benadering voor het bestaande leesonderwijs voorstellen, waarbij we actuele internationale inzichten rondom tekstbegrip een plaats willen geven. Het boek bestaat dus deels uit beschrijvend onderzoek, met een duidelijk evaluerende component, en daarnaast uit ontwerp- en interventieonderzoek.

In het eerste beschrijvende deel (hoofdstuk 2 tot en met 4) speelt de tekst-metvragen-didactiek een belangrijke rol. Tot in de jaren negentig van de vorige eeuw

propageerden vakdidactische publicaties rondom het schoolvak Nederlands deze didactiek: het beantwoorden van tekst(on)afhankelijke begripsvragen bij een tekst werd beschouwd als een geschikt middel om tekstbegrip te kweken (bijv. Drop, 1979; Griffioen, 1975; Kreeft & Luijten, 1979). In deze toenmalige studies speelden twee niet nader onderbouwde aannames een cruciale rol. Ten eerste: wie een tekst goed voorafleest, zou vervolgens ook begripsvragen goed maken (Drop, 1979, p. 17). En ten tweede: doordat leerlingen een systematiek krijgen aangereikt om een tekst te bestuderen, zou het beantwoorden van begripsvragen leerlingen stimuleren “zelfstandig teksten te leren lezen” (bijv. Hendrix & Hulshof, 1994, p. 31). Op deze opvattingen is sinds eind jaren negentig in toenemende mate (felle) kritiek geuit vanuit vakdidactici, wetenschappers en docenten Nederlands: de ‘tekst met vragen’ kan misschien een valide leestoets zijn, dat maakt haar nog niet geschikt als trainingsinstrument (bijv. Hoogeveen & Bonset, 1998; Westhoff, 2012). Hoe kritisch de tekst met vragen als leerinstrument vandaag de dag ook vaak bejegend wordt, een empirische onderbouwing bij de kritiek ontbrak grotendeels. En zoals we al vaststelden, bepaalt de tekst-met-vragen-didactiek nog altijd vaak de leesles Nederlands in het voortgezet onderwijs (bijv. Frank et al., 2019).

In het beschrijvend-evaluerende deel van dit boek (hoofdstuk 2 tot en met 4) staan daarom de volgende twee onderzoeksvragen centraal:

- a) Hoe bouwen vwo-leerlingen tekstbegrip op en welke ontwikkeling maken ze hierin door, als we uitgaan van de reguliere praktijk in de lessen tekstbegrip in het schoolvak Nederlands?
- b) In hoeverre vormt de tekst met vragen een effectieve taak om tekstbegrip te oefenen in de lessen tekstbegrip in het schoolvak Nederlands?

In het tweede deel, het ontwerp- en interventiedeel van dit boek (hoofdstuk 5 en 6), willen we een alternatieve, meer eigentijdse didactiek voor de lessen tekstbegrip in de vwo bovenbouw ontwerpen. De laatste twee decennia staan veel studies rondom het schoolvak Nederlands een meer procesgerichte didactische benadering voor van taalvaardigheidsonderwijs (bijv. Braaksma, 2002; Couzijn, 1995; Rijlaarsdam et al., 2008), waarbij leerlingen eerst op hun eigen denkproces of andermans denkproces bij het uitvoeren van een taak reflecteren, voordat ze met een taak aan de slag gaan; de reguliere tekst-met-vragen-didactiek richt zich daarentegen vooral op een eenduidig te beoordelen leesproduct: de al dan niet juiste antwoorden van leerlingen op begripsvragen (bijv. Hoogeveen & Bonset, 1998; Van de Wetering & Groenendijk, 2015). Uitgaande van een meer procesgerichte benadering en van recente wetenschappelijke inzichten over tekstbegrip, willen we een alternatieve didactiek voor de les tekstbegrip ontwikkelen en beproeven. In dit ontwerpdeel van dit boek staat daarom de volgende onderzoeksvraag centraal:

- c) Wat is een effectieve procesgerichte didactiek voor de lessen tekstbegrip in het schoolvak Nederlands?

2.3 *Uitgangspunten bij het onderzoek*

Dit boek is een bundeling van vijf artikelen, waarvan vier er in een eerdere versie een plek vonden in Nederlandse wetenschappelijke tijdschriften: één verscheen als artikel in *Levende Talen Tijdschrift*; drie verschenen als artikel in *Pedagogische Studiën*, waarvan er één een prijs mocht ontvangen.⁴ Eén (hoofdstuk 6) staat bij het perse gaan van dit boek op het punt van verschijnen. Voor een aantal alinea's in deze inleiding en in het afsluitende discussiehoofdstuk is daarnaast geput uit twee artikelen uit *Levende Talen Magazine*, waarin we stilstonden bij pijnpunten in de huidige leesdidactiek in het schoolvak Nederlands (Rooijackers et al., 2021a, 2021b).

Deze opzet kent zeker haar beperkingen: wie alle hoofdstukken aandachtig doorneemt, kan bijvoorbeeld opmerken dat hoofdstuk 5 nogal geserreerd is in vergelijking met vooral de hoofdstukken 3, 4 en 6. De oorzaak is eenvoudig: hoofdstuk 5 verscheen oorspronkelijk in *Levende Talen Tijdschrift*, dat in vergelijking met *Pedagogische Studiën* een beperktere woordlimiet hanteert. Door de artikelenopzet van dit boek is er bovendien een duidelijke overlap en herhaling waarneembaar in hoofdstuk 2 tot en met 6. Dat is misschien onprettig voor wie het boek als geheel leest, het voordeel is dat elk hoofdstuk op zichzelf leesbaar is.

Al bereikt een artikelenbundel misschien minder gemakkelijk logische en retorische samenhang dan een monografische studie, daarmee is dit nog geen lukrake verzameling van publicaties rondom tekstbegrip. Een zevental onderling sterk samenhangende uitgangspunten schragen deze vijf studies:

- I. diep begrip centraal
- II. vaardige lezers centraal
- III. tekstbegrip in beeld (en in geluid)
- IV. leesproces versus antwoordproces
- V. doelmatig strategiegebruik
- VI. ecologische validiteit
- VII. generaliseerbaarheid

We lopen hier kort deze uitgangspunten langs, waarbij we telkens aangeven hoe een uitgangspunt zich vertaalde in enerzijds het beschrijvende deel van dit boek en anderzijds het ontwerpgedeelte van dit boek.

I. Diep begrip centraal. Het opbouwen van begrip bij een tekst is een complex samenspel van processen: er spelen allerlei lagere-orde-, *bottom-up*-processen een rol, zoals het decoderen van woorden, het toekennen van betekenissen aan woorden, het syntactisch integreren van woorden tot zinsdelen en het construeren van lokale samenhang tussen zinsdelen (Perfetti & Stafura, 2014). Aan die lagere-orde-processen zullen we in dit boek weinig woorden wijden. Het gaat in dit boek om hogere-orde-begripsprocessen: het construeren van globale samenhang tussen zinnen, tussen alinea's en tussen blokken van alinea's. Daarbij maken we steeds gebruik van het begrip

⁴ De tijdschriftversie van de studie in hoofdstuk 4 werd door de VOR, VFO en de redactie van *Pedagogische Studiën* onderscheiden als het beste artikel van 2021 in dat tijdschrift.

situatiemodel zoals uitgewerkt in de Constructie-Integratie-theorie (Kintsch, 1998): we richten ons niet op het begrip waarbij lezers *verbatim* tekstbeweringen opslaan en onthouden – in een ‘tekstbasis’ –, maar op het begrip waarbij lezers de beweringen in de tekst hebben aangevuld met én geïntegreerd in hun kennis van de wereld – in een ‘situatiemodel’. In het eerste beschrijvende deel, hoofdstuk 2 tot en met 4, willen we daarom vooral beschrijven in hoeverre vwo-leerlingen tijdens het vooraflezen van teksten en tijdens het beantwoorden van begripsvragen situatiemodelbegrip opbouwen. We zullen daarvoor (net als in het tweede ontwerpgedeelte van het boek) steeds leestaken gebruiken die vooral een appel doen op situatiemodelbegrip.

In het tweede deel van het boek, het ontwerpgedeelte (hoofdstuk 5 en 6), richten we ons vanuit dit situatiemodelbegrip op het sterk hieraan verbonden concept ‘diep lezen’: het “geconcentreerd lezen, waarbij de samenhang en betekenis van een tekst worden ervaren” (Onderwijsraad & Raad voor Cultuur, 2019). Diep lezen omvat dan “the array of sophisticated processes that propel comprehension and that include inferential and deductive reasoning, analogical skills, critical analysis, reflection, and insight” (Wolf & Barzillai, 2009, p. 32; vgl. Baron, 2015; Hayles, 2012; Piper, 2012; Wolf, 2018). In de lessenserie willen we leerlingen leren wat diep begrip is, hoe ze het kunnen opbouwen en wanneer en waarom ze het moeten inzetten.

II. Vaardige lezers centraal. Vwo-(bovenbouw)leerlingen vormen de vaardigere lezers in het Nederlandse voortgezet onderwijs (Expertgroep Doorlopende Leerlijnen Taal, 2008, 2009; Gubbels et al., 2019, Van Silfhout, 2014). Er is zeker reden om specifiek deze groep te belichten. In het Referentiekader Taal wordt de beschrijving van het beheersingsniveau van vwo-abituriënten, niveau 4F, nauwelijks tot niet ondersteund door empirisch onderzoek, in tegenstelling tot de niveaubeschrijvingen 1F en 2F (Expertgroep Doorlopende Leerlijnen Taal, 2008); in een interview spraken de voorzitters van de toenmalige Expertgroep daarom over 4F als “een educated guess” (Robben, 2008). Er zijn daarnaast zorgen over het tekstbegrip van instromende leerlingen op Nederlandstalige universiteiten (Bonset, 2010; Van Houtven et al., 2010; Raad voor Nederlandse Taal en Letteren, 2015; Warps et al., 2021); deze zouden over het algemeen te weinig kritisch kunnen lezen (Warps et al., 2021). En het zijn juist deze academisch georiënteerde leerlingen die later in studie en werk vaak geconcentreerd en diepgaand een tekst zullen moeten bestuderen en die daarbij in kwaliteit wisselende informatiebronnen over een onderwerp zullen moeten evalueren en integreren (Goldman, 1997; Hermida, 2009; Perfetti, 1997).

De focus op deze leerlingen staat ons toe ons vooral te richten op diep begrip; juist deze leerlingen zouden over het algemeen weinig tot geen moeite moeten hebben met bottom-up-begripsprocessen (Expertgroep Doorlopende Leerlijnen Taal, 2008, 2009). In het beschrijvend deel (hoofdstuk 2 tot en met 4) kunnen we daarom bij de tekst met vragen de relatie onderzoeken tussen het voorafbestuderen van een tekst en het beantwoorden van vragen: als deze relatief vaardige leerlingen een tekst vooraf diepgaand hebben bestudeerd (en dus een relatief hoogwaardig begrip hebben opgebouwd), halen ze dan een hogere score op begripsvragen, dan als ze dezelfde tekst vooraf nogal oppervlakkig hebben doorge-

nomen? In het interventiedeel (hoofdstuk 5 en 6) staat dit uitgangspunt ons toe vakinhoudelijk en -didactisch de meer basale begripsvaardigheden over te slaan en juist vooral de opbouw van diep begrip centraal te stellen.

III. Tekstbegrip in beeld (en in geluid). Tekstbegrip is een fenomeen dat een beschouwer in eerste instantie alleen indirect kan waarnemen; alleen vanuit symptomen, artefacten, residuen leiden we af in hoeverre mensen begrip bij een tekst opbouwen (bijv. Pearson & Cervetti, 2013). In het huidige leesonderwijs in het schoolvak Nederlands doet men dat, zoals we zagen, vaak via antwoorden op begripsvragen (bijv. Hoogeveen & Bonset, 1998). De relatie van dit leesproduct met het begripsproces is echter complex: bij het beantwoorden van begripsvragen kunnen tal van vaardigheden een rol spelen die niet direct op tekstbegrip betrekking hebben, zoals schrijfvaardigheid en *test-wiseness* (bijv. Cohen & Upton, 2006; Rupp et al., 2006). In dit boek proberen we tekstbegrip daarom dichter op de huid te zitten, vooral door lees- en antwoordprocessen met moderne onderzoekstechnieken te visualiseren en hoorbaar te maken.

In het beschrijvende deel (hoofdstuk 2 tot en met 4) doen we dit door de oogbewegingen van lezende leerlingen te volgen via een eye-tracker. Voordeel van eye-tracking is dat het het leesgedrag feitelijk weergeeft; nadeel ervan is dat het geen directe uitspraken toestaat over het doel achter dit leesgedrag, een gegeven waarin hardopdenkmethoden juist wel kunnen voorzien (Holmqvist & Andersson, 2017). Door leerlingen tijdens een leestaak te eye-tracken én door hen naderhand te laten reflecteren op het doel van hun leesgedrag, zullen we de kracht van eye-tracking en hardopdenkmethoden bundelen (vgl. Bax, 2013; Brunfaut & McCray, 2015). Zo kunnen we directer dan via een leestoets uitspraken doen over de kwaliteit van het begripsproces bij leerlingen én kunnen we de relatie tussen leesproces en leesproduct nader verkennen.

In het ontwerpdeel (hoofdstuk 5 en 6) zetten we visualisatie van tekstbegrip in als een didactisch instrument: we gebruiken eerder gemaakte oogbewegingsfilmpjes en audio-opnamen van lezende leeftijdsgenoten om leerlingen klassikaal te laten reflecteren op tekstbegrip en hen dat te laten evalueren: leerlingen kunnen nauwkeurig op een beeldscherm volgen wat leeftijdsgenoten in een tekst lezen en ze luisteren naar de bedoelingen van deze leeftijdsgenoten met hun leesgedrag. Daarmee sluiten we aan bij onderzoek naar de effectiviteit van *Eye-Movements Modeling Examples* (EMME's: zie Van Marlen, 2019) in het onderwijs.

IV. Leesproces versus antwoordproces. In de bestaande praktijk in het schoolvak Nederlands worden vaak uitspraken gedaan over de leesvaardigheid van leerlingen via producten, de antwoorden op begripsvragen bijvoorbeeld; hun leesgedrag zelf blijft vaak buiten beschouwing (bijv. Linthorst & De Glopper, 2015). In dit boek richten we ons nadrukkelijk op begripsprocessen: we onderzoeken zowel het leesproces als het antwoordproces van leerlingen bij leestaken en we onderzoeken de relatie van dit leesproces met dit antwoordproces en met het leesproduct dat leerlingen vervolgens opleveren (i.e. hun antwoorden op begripsvragen). Daarmee sluiten we aan bij een meer procesgerichte benadering van taalvaardigheid binnen het schoolvak Nederlands (bijv. Braaksma, 2002; Couzijn, 1995; Rijlaarsdam et al., 2008).

In het beschrijvende deel (hoofdstuk 2 tot en met 4) richten we ons vanuit dit uitgangspunt op het lees- en antwoordgedrag van leerlingen zelf. In hoofdstuk 2 en 3 zullen we zowel het vooraf-lees- als het antwoordproces van vwo-leerlingen bij teksten met vragen via eye-tracking afzonderlijk registreren, beschrijven en analyseren. Daarbij onderzoeken we óók de relatie tussen vooraf-leesproces en antwoordproces: kunnen we aantonen dat leerlingen die een tekst vooraf goed bestuderen, ook vaker de begripsvragen goed beantwoorden? Om die vraag scherper te kunnen beantwoorden, zullen we in hoofdstuk 4 de relatie tussen vooraf-leesproces en het antwoordproces experimenteel manipuleren.

In het ontwerp- en interventiedeel (hoofdstuk 5 en 6) richten we ons op het leesproces van leerlingen via observerend leren lezen. De meerwaarde van observerend leren voor het domein schrijfvaardigheid is inmiddels vrij breed aangetoond (bijv. Bonset & Braaksma, 2008; Braaksma, 2002; Couzijn, 1995; Rijlaarsdam et al., 2008); studies die de meerwaarde hiervan voor de *leesles* Nederlands laten zien, zijn aanmerkelijk schaarser (Couzijn, 1995; Keehnen et al., 2015), terwijl de werkwijze in de leesles Nederlands, zoals gezegd, vaak nog bij uitstek productgericht is (bijv. Linthorst & De Gloppe, 2015; vgl. Bogaerds-Hazenbergh et al., 2022b).

V. Doelmatig strategiegebruik. In de leesles Nederlands in het voortgezet onderwijs maakt men momenteel vaak gebruik van een weinig evidencebased invulling van strategie-instructie (vgl. Bonset et al., 2019; zie Rooijackers et al., 2021b). Zoals we in § 1.3 zagen, baseren leergangen Nederlands zich hierbij op Drop en De Vries (1974) die de stappen uit een Amerikaanse studeermethode uit 1946 hebben omgebouwd tot een set ‘strategieën’; deze zijn echter vooral als generieke manieren van lezen te beschouwen. In dit boek willen we ons nadrukkelijk wél baseren op evidencebased leestategieën (Pressley & Afflerbach, 1995): we beschouwen strategieën als “deliberate, goal-directed attempts to control and modify the reader’s effort to decode tekst, understand words, and construct meanings of text”, zoals een veelgebruikte definitie luidt (Afflerbach et al., 2008, p. 368). Kernachtig gezegd: strategieën zijn middelen die een lezer bewust inzet om een bepaald leesdoel te bereiken. Niet voor niets is het vooropstellen en in ogenschouw houden van het leesdoel tijdens lezen een bij uitstek krachtige component in strategie-instructie (Okkinga et al., 2018). De vraag in hoeverre lezers via strategieën een specifiek leesdoel bereiken, zullen we daarom steeds als vertrekpunt nemen van onderzoek.

Daarbij vertrekken we vanuit de opvatting dat strategiegebruik sterk persoons-, context- en doelafhankelijk is; we zullen, met andere woorden, een flexibele opvatting van strategie-instructie hanteren (Aarnoutse & Verhoeven, 2002). Welke begripsproblemen lezers ervaren tijdens lezen, is namelijk deels individueel bepaald; vaardige lezers ervaren andere leesproblemen bij een complexe tekst dan minder vaardige lezers en zetten daarmee strategieën vaak sterk onderling afwijkend in (bijv. Pressley & Afflerbach, 1995; Walczyk & Griffith-Ross, 2007). Bovendien kan de inzet van strategieën sterk variëren in gebruikcontexten (Afflerbach et al., 2020). In het Nederlandse leesonderwijs lijkt deze contextgebondenheid, het

conditionele karakter van strategiegebruik echter nogal eens buiten beschouwing te worden gelaten (Bogaerds-Hazenberg et al., 2022b).

In hoofdstuk 2 en 3 beschrijven we in het vooraf-leesproces bij teksten met vragen de mate waarin vwo-leerlingen een belangrijke leesstrategie van vaardige lezers inzetten: het terugkijken naar kernzinnen (Pressley & Afflerbach, 1995). Daarmee stellen we onszelf de vraag in hoeverre leerlingen het beoogde doel bereiken: diep begrip van een tekst. En we stellen onszelf de vraag in hoeverre het al dan niet toepassen van deze strategie tijdens vooraf-lezen een ander doel kan dienen: een hogere score behalen bij het beantwoorden van begripsvragen.

In het ontwerp- en interventiegedeelte (hoofdstuk 5 en 6) willen we leerlingen aan het denken zetten over de vraag in hoeverre leeftijdsgenoten via hun strategisch leesgedrag tot diep begrip bij een tekst komen en, via deze weg, ze aan het denken zetten over de effectiviteit van hun eigen strategisch leesgedrag. Zo willen we leerlingen stimuleren tot meer ‘diep-leesgedrag’.

VI. Ecologische validiteit. We zullen in dit boek zoveel mogelijk een natuurlijke onderzoeksomgeving nastreven. Op de studies in hoofdstukken 2 en 3 na, die noodzakelijkerwijs plaatsvonden onder min of meer experimentele omstandigheden (namelijk in een kamer met een eye-tracker op een middelbare school), zullen we de studies steeds inbedden in reguliere toets- (hoofdstuk 4) of lessituaties op Nederlandse middelbare scholen (hoofdstuk 5 en 6). Daarbij waren op allerlei niveaus leraren Nederlands betrokken: als screeners, als beoordelaars, als organisatoren en als docenten van experimenteel lesmateriaal.

Daarbij willen we in dit boek aansluiten op de bestaande praktijk in (vooral de bovenbouw vwo binnen) het schoolvak Nederlands. Dat geldt voor de tekstkeuze, opdrachtkeuze en context waarbinnen de studies zich situeren. Overeenkomstig de praktijk in de leesparagrafen van huidige courante leergangen Nederlands, kiezen we steeds zakelijke teksten uit dagbladen en tijdschriften (in hoofdstuk 2, 3, 5 en 6) of selecteerden we zakelijke teksten met vragen uit het Centraal Examen Nederlands havo/vwo (hoofdstuk 4): met de teksten-met-vragen-sets uit deze omstandig geconstrueerde toets wordt namelijk vaak geoefend in de leesles in de bovenbouw havo/vwo (Canton et al., 2013).

Door onderzoek zoveel mogelijk in te bedden in de bestaande lespraktijk en door een actieve participatie van (collega-)docenten Nederlands na te streven, hopen we de ecologische validiteit van ons onderzoek te vergroten.

VII. Generaliseerbaarheid. We eindigen deze bespreking van de uitgangspunten met een meer impliciet uitgangspunt: de generaliseerbaarheid van de bevindingen in onze studies. In alle hoofdstukken zullen uitkomsten via multiniveau-modellen worden geanalyseerd, wat ons zal toestaan te generaliseren over leerlingen en taken, en in mindere mate teksten. Daarnaast zullen we uitkomsten uit studies trachten te repliceren. Zo zullen we niet alleen in hoofdstuk 2 en 3 bij teksten met vragen de relatie tussen vooraf-lezen en de antwoordscore beschrijven, maar we zullen deze relatie in hoofdstuk 4 vervolgens ook manipuleren. En de effectiviteit

van een experimentele lessenserie onderzoeken we allereerst in hoofdstuk 5, waarin docenten met hulp van onderzoekers een eerste versie van de lessenserie afnemen, en deze effectiviteit onderzoeken we daarna opnieuw in hoofdstuk 6, waarin docenten zonder hulp van onderzoekers de definitieve versie van de lessenserie afnemen.

2.4 *Algemene opzet*

Hoe is dit boek opgebouwd? Het vervolg van dit boek bestaat uit vijf hoofdstukken én een afsluitend discussiehoofdstuk. De eerste helft van het boek is, zoals gezegd, vooral beschrijvend. In hoofdstuk 2, een *mixed-method* eye-tracking-studie, wordt allereerst het lees- en antwoordgedrag van vwo 4-leerlingen beschreven, bij vier teksten met vragen. In hoofdstuk 3 beschrijven we vanuit dezelfde onderzoeksmethode verschillen in lees- en antwoordgedrag tussen drie groepen leerlingen, vwo 2-, vwo 4- en vwo 6-leerlingen. In beide hoofdstukken zal de aandacht ook uitgaan naar de relatie tussen het vooraf-lezen van teksten en het beantwoorden van vragen.

In hoofdstuk 4, een quasi-experimentele studie, onderzoeken we de tekst met vragen als instrument om tekstbegrip te trainen. We bestuderen de relatie tussen het vooraf-lezen van de tekst en het beantwoorden van de vragen, vanuit twee vragen: Als leerlingen bij deze taak bij de beantwoording van de vragen de tekst niet meer kunnen raadplegen, hoe beïnvloedt dat dan de antwoordscore? En hoe wordt de antwoordscore beïnvloed als we leerlingen dwingen een tekst vooraf intensief te bestuderen, voordat ze de vragen kunnen beantwoorden?

In de tweede helft van dit boek, het ontwerp- en interventiedeel, zullen we vervolgens inzetten op het ontwerpen en testen van een alternatieve inhoud en didactiek in tekstbegrip, waarbij we vooral zullen aansluiten bij de leerprincipes van observerend leren (Bandura & Walter, 1977; Bandura, 1986) en bij recente theoretische inzichten rondom tekstbegrip (bijv. Britt et al., 2018). Hoofdstuk 5 en 6 bevatten twee interventiestudies: hoofdstuk 5 is een pilot-studie, waarin een eerste versie van de lessenserie kleinschalig, klassikaal, onder leiding van een docent Nederlands en een onderzoeker, met vooral papieren middelen wordt afgenomen (hoofdstuk 5); hoofdstuk 6 bevat een uitgebreidere interventiestudie waarin een definitieve versie van de lessenserie grootschalig, digitaal en voor een belangrijk deel individueel wordt afgenomen, waarbij de docent Nederlands in één conditie wel en in één conditie niet als begeleider optreedt (hoofdstuk 6).

In het slothoofdstuk, de discussie, staan we ten slotte stil bij de algemene bevindingen in dit onderzoek: wat zijn de belangrijkste bevindingen uit hoofdstuk 2 tot en met 6, hoe luiden de antwoorden op de algemene onderzoeksvragen uit § 2.2, en wat zijn de implicaties van de bevindingen voor de wetenschap en voor de lessen tekstbegrip in het schoolvak Nederlands, specifiek voor vwo-leerlingen?

2.5 *Enkele opmerkingen over de ontstaansgeschiedenis*

Voordat we ons eigenlijke onderzoek beschrijven in hoofdstuk 2 tot en met 6, merken we kort nog vier zaken op over de ontstaansachtergronden van dit boek.

Allereerst: zoals in een meerjarig project als ons onderzoek te verwachten is, is er in dit boek sprake van voortschrijdend inzicht. In hoofdstuk 2 en 3 beschrijven we zoveel mogelijk waardenvrij het vooraf-lees- en antwoordgedrag van Nederlandse vwo-leerlingen bij vier teksten met vragen. De resultaten daarvan waren reden om een experiment te ontwerpen waarmee we – meer dan in de oorspronkelijke onderzoeksplannen beoogd was – kritisch het functioneren van de reguliere tekst-met-vragen-oefening konden onderzoeken: hierover zullen we rapporteren in hoofdstuk 4. Daarmee vormt ook de experimentele lessenserie in met name hoofdstuk 6 – meer dan in de oorspronkelijke onderzoeksplannen beoogd was – een tegenhanger van de tekst-met-vragen-didactiek. Dat in dit boek de tekst-met-vragen-aanpak gaandeweg kritisch wordt benaderd, is resultaat van ons onderzoek, niet een vooropgezette bedoeling.

Ten tweede: de Covid19-pandemie heeft onmiskenbaar de inhoud van ons onderzoek beïnvloed. De experimenten in hoofdstuk 4, 5 en 6 zijn afgenomen tussen twee maatschappelijke lockdowns door, in twee schooljaren waarin de schooldeuren soms landelijk maandenlang gesloten waren. Op de lastige vraag in hoeverre deze bijzondere onderwijssituatie de resultaten in ons onderzoek heeft beïnvloed, zullen we in elk van de hoofdstukken 4, 5 en 6 stilstaan; in het algemeen lijkt de motivatie, concentratie en discipline van Nederlandse leerlingen tijdens de epidemie te zijn aangetast (NJI, 2022). De *inhoud* van ons onderzoek is echter zeker deels erdoor veranderd. In het onderzoek in hoofdstuk 6 is de lessenserie bijvoorbeeld geheel gedigitaliseerd en leerlingen zullen deze lessenserie in een van de twee onderzoekscondities geheel zelfstandig doorwerken. Deze beide zaken waren een direct gevolg van de ongewisse situatie in het Nederlandse onderwijs: een algehele schoolsluiting hing in de jaren 2020 en 2021 regelmatig in de lucht. Door ons experiment op deze manier in te richten, leek een afname ervan binnen deze onzekere periode gewaarborgd.

Ten derde: het is goed op te merken dat het onderzoek ethisch de toets der kritiek heeft doorstaan. Voor alle experimenten in dit boek geldt dat ze uitgevoerd zijn met toestemming van de Facultaire Ethische Toetsingscommissie Geesteswetenschappen (FETC-GW) van de Universiteit Utrecht. En alle experimenten in ons onderzoek zijn met instemming van scholen, docenten en leerlingen afgenomen.

Tot besluit nog een opmerking over de taal in dit boek: dit boek is, op een Engelstalige samenvatting na, geschreven in het Nederlands. Dat is in universitaire kringen in het huidige tijdsgewricht bijna een stille daad van verzet, en zo mag deze keuze gezien worden. De voornaamste reden achter de keuze was echter een andere: dit boek houdt zich bezig met het schoolvak Nederlands, wat lag er meer voor de hand dan het in de Nederlandse taal zelf te schrijven, opdat het zijn weg moge vinden naar het klaslokaal Nederlands? ... Los van andere, meer nostalgische redenen, die soms des avonds opkomen voordat men slapen gaat – want, ach,

Mijne moedertaal is de schoonste taal,
 En zou ze 't ook niet wezen,
 Haar verkiezen zou ik nog,
 Want zij is de mijne toch! (De Cort, 1868, p. 119)

HOOFDSTUK 2

Teksten lezen, vragen beantwoorden,
geobserveerd bij vwo 4-leerlingen

*Een oogbewegingsonderzoek**

* Dit hoofdstuk is in een eerdere vorm verschenen als: Rooijackers, P., Silfhout, G. van, Schuurs, U., Mulders, I., & Bergh, H. van den. (2020). Lezen en antwoorden bij de tekst met vragen geobserveerd; Een eye-trackstudie onder vwo 4-leerlingen. *Pedagogische Studiën*, 97(1), 42-50.

Samenvatting

In leeslessen Nederlands in het secundair onderwijs wordt vaak de tekst met vragen gebruikt om leerlingen te toetsen en trainen in tekstbegrip. Daarbij is de gedachte doorgaans: wie goed vooraf leest, maakt de vragen ook goed.

In deze studie wordt de relatie tussen het vooraf-lezen van teksten en het beantwoorden van bijbehorende vragen onderzocht door analyse van de oogbewegingen. Vertrekpunt vormt de aanname in de Constructie-Integratietheorie van Kintsch (1998) dat lezers voor een dieper tekstbegrip meer aandacht besteden aan kernzinnen dan aan niet-kernzinnen.

Er is onderzocht in hoeverre vwo 4-leerlingen ($N = 16$) tijdens vooraf-lezen meer leestijd schenken aan kernzinnen dan aan niet-kernzinnen en in hoeverre dit verband houdt met het antwoorden. Uit de resultaten blijkt dat deelnemers tijdens vooraf-lezen meer tijd besteden aan kernzinnen, maar op dieper tekstbegripsniveau kon geen relatie tussen vooraf-lezen en antwoorden worden vastgesteld. Deelnemers lezen vooraf over het algemeen lineair en construeren pas tijdens het antwoordproces een dieper tekstbegrip. De vraag is daarmee in hoeverre de huidige inrichting van deze taak tekstbegrip bevordert en de dominantie ervan in het schoolvak Nederlands gerechtvaardigd is.

1 Inleiding

1.1 *Het tekstbegrip van vwo-leerlingen*

Het lezen van zakelijke teksten in de hoogste jaarlagen van het Nederlandse vwo is weinig bestudeerd. In 2010 stelde de Expertgroep Doorlopende Leerlijnen Taal normen op voor zakelijk lezen voor vwo 6 (in Vlaanderen het laatste jaar van de derde graad in het aso). De beschrijving van dit zogenaamde ‘niveau 4F’ kon echter nauwelijks worden ondersteund door empirische studies. De voorzitters van de Expertgroep spraken daarom in 2008 al van een “educated guess” (Robben, 2008).

De Expertgroep Doorlopende Leerlijnen Taal stelde ook voor havo 5 normen op voor zakelijk lezen: ‘niveau 3F’, dat grofweg ook voor vwo 4-leerlingen geldt (het laatste jaar van de tweede graad in het aso). Leerlingen zouden vanaf dit niveau een grote variatie aan teksten “over opleidings- en beroepsgebonden en maatschappelijke onderwerpen” kunnen lezen, met “begrip voor geheel en details”. Een nadere beschrijving van niveau 3F en 4F is te vinden in Expertgroep Doorlopende Leerlijnen Taal (2008, 2009).

De vraag is of we op dit moment enig inzicht hebben in het werkelijke tekstbegrip van de Nederlandse vwo-bovenbouwleerlingen. Belangrijkste dataverzameling over de vwo-lezer vormt het Nederlandse Centraal Schriftelijk Examen vwo waarvan we kunnen aannemen dat dit tekstbegrip toetst op niveau 4F (Van Onna et al., 2014): vwo-abituriënten zouden dus teksten kunnen begrijpen op niveau 4F. Onder deze laatste conclusie ligt de vaak in het schoolvak Nederlands gemaakte aanname dat wie goed vooraf heeft gelezen, ook de erop volgende vragen goed beantwoordt. Echter, de relatie tussen vooraf-lezen en beantwoorden ligt in werkelijkheid aanmerkelijk complexer: we weten dat een vooraf gestelde taak zelf een sterke invloed heeft op het lezen (bijv. André, 1979; Cerdán et al., 2009; Hamilton, 1985; Yeari et al., 2015). In internationaal onderzoek krijgen kandidaten dan ook meestal vooraf een taakcontext aangeboden voordat ze een tekst gaan lezen (Pirls, PISA, IELTS etc.). Strikt genomen zeggen de resultaten van het Centraal Examen Nederlands vwo alleen iets over of leerlingen op niveau 4F antwoorden bij een ‘tekst met vragen’. Over het tekstbegrip tijdens het lezen vóórdat ze de vragen beantwoordden, weten we vrijwel niets.

Om in deze twee kwesties – het tekstbegrip van de vwo bovenbouw-leerling en de relatie tussen vooraf-lezen en beantwoorden bij de tekst met vragen – meer

duidelijkheid te brengen, onderzoeken we in deze studie de relatie tussen het vooraf-lezen en het beantwoorden. We volgden daarom de oogbewegingen van Nederlandse vwo bovenbouw-leerlingen terwijl ze korte zakelijke teksten lazen én terwijl zij bijbehorende vragen over de tekst beantwoordden.

1.2 Relatie tussen vooraf-lezen en vragen beantwoorden

Het leesgedrag van een expert-lezer is redelijk scherp in kaart gebracht. Bepalend is met name de studie van Pressley en Afflerbach (1995) waarin 38 hardopdenkstudies zijn geanalyseerd om de algemene kenmerken van een volwassen lezer op te kunnen stellen. Het blijkt dat een goede lezer zich vooral kenmerkt door de inzet van tekstbegrip bevorderende leesstrategieën, zoals leesdoelen stellen en voorkennis activeren tijdens het lezen. Deze bevindingen hebben inmiddels hun weg gevonden in tal van overzichtsstudies over tekstbegrip (Duke & Pearson, 2002; McNamara, 2007; Perfetti et al., 2005, etc.).

Het natuurlijke leesproces is lastig in kaart te brengen. In de jaren tachtig en negentig maakten onderzoekers vooral gebruik van hardopdenkprotocollen om zicht te krijgen op het verloop van het leesproces en het rationale erachter (vgl. Pressley & Afflerbach, 1995). Maar belangrijk nadeel van hardopdenkprotocollen is dat de verbalisatie tijdens het lezen het natuurlijke handelingsproces danig verstoort (Gerjets et al., 2011).

Vanaf de jaren negentig wordt onder andere daarom oogbewegingenregistratie steeds vaker in onderzoek ingezet om het lezen te registreren, al richten deze studies zich wel vaak op deelaspecten van het lezen (bijv. Cozijn, 2000; Van Silfhout, 2014; Yeari et al., 2015). Eye-tracking registreert waarop en hoelang ogen fixeren. De methode gaat ervan uit dat verschillen in de duur van fixaties verschillen in cognitieve verwerkingstijd vertegenwoordigen; bij ‘saccades’, oogverplaatsingen, wordt de aandacht verschoven (Duchowski, 2007; Holmqvist & Andersson, 2017). Eye-tracking verstoort het natuurlijke leesproces minder dan hardop denken en kent bovendien inmiddels weinig technische beperkingen. Eye-tracking-data verstrekken evenwel geen rationale bij het waargenomen leesgedrag (Bax, 2013; Cozijn, 2000).

Het is tegenwoordig duidelijk dat vooraf-lezen niet eenduidig de kwaliteit van een leesproduct bepaalt. Een taak die vooraf wordt gegeven, heeft zelf juist een belangrijk effect op het lezen (bijv. André, 1979; Cerdán et al., 2009; Hamilton, 1985; Lorch et al., 1987; Narvaez et al., 1999). Een taak met open vragen zet lezers bijvoorbeeld meer aan tot diepgaand begrip dan een taak met gesloten vragen (Magliano et al., 2007; Ozuru et al., 2013). En Yeari et al. (2015) vonden dat lezers bij een samenvattingsopdracht meer tijd besteedden aan hoofdzaken, dan bij een taak met gesloten vragen.

De tekst met vragen lijkt het zakelijk lezen in het klaslokaal Nederlands in de vwo bovenbouw te domineren (Meestringa & Ravesloot, 2012; Canton et al., 2013). Deze leestaak en -toets kent in het schoolvak Nederlands een lange traditie (Hulshof et al., 2015) en huidige Nederlandse bovenbouwmethodes als *Nieuw Neder-*

lands en *Op niveau* leggen er een sterk accent op (vgl. Frank et al., 2019; Merckx et al., 2018). Ook het landelijk Centraal Schriftelijk Examen Nederlands vwo vertrekt vanuit deze taak en heeft daarmee naar alle waarschijnlijkheid een sterk terugslageffect op de inrichting van het leesonderwijs in het schoolvak Nederlands (Meestringa & Ravesloot, 2012). De grondslag van deze taak is meestal dezelfde. Leerlingen moeten, zonder uitgebreide taakcontext vooraf, een tekst grondig bestuderen, met oog voor hoofdzaken, hoofdgedachte, structuur en alineafuncties; daarna volgt een aantal vragen, en de tekst blijft dan steeds raadpleegbaar.

Er is in Nederland momenteel een duidelijke onvrede over deze situatie: de tekst met vragen zou leiden tot een oppervlakkige bestudering van de tekst vooraf en ze zou zelf centraal zijn komen te staan in didactiek en vakinhoud (SLO, 2016a; SLO, 2017; Nederlands Nu! & Sectiebestuur Nederlands LT, 2018; zie eerder al Hoogeveen & Bonset, 1998). Deze onvrede bereikte in Nederland de afgelopen jaren herhaaldelijk de media (bijv. De Bruin, 2018). De empirische onderbouwing bij dit standpunt is beperkt (SLO, 2017), al zijn er een aantal studies waaruit blijkt dat een score bij de tekst met vragen slechts in beperkte mate het diepere begrip van een tekst weerspiegelt (Kamalski, 2007; Land, 2009).

In deze studie willen we de relatie tussen het vooraf-lezen van een tekst en het beantwoorden van vragen verkennen door middel van eye-tracking, dat het natuurlijke lezen immers dicht benadert.

Om de kwaliteit van het vooraf-lezen en beantwoorden te beschrijven, gaan we uit van de Constructie-Integratietheorie van Kintsch (1998), vertrekpunt van vele Nederlandse en internationale leesonderzoeken (vgl. McNamara & Magliano, 2009a). Volgens Kintsch (1998) is een tekst een netwerk van hiërarchisch samenhangende kleine, betekenisvolle beweringen (proposities). Een lezer is steeds op zoek naar deze samenhang en construeert tussen woorden en zinnen verbindingen (een microstructuur) en tussen alinea's (een macrostructuur). Dit constructieproces mondt dan volgens Kintsch uit in een 'tekstbasis', een vrij letterlijke mentale representatie van de tekst, en in een 'situatiemodel', een model waarin de lezer de wereld van de tekst opneemt in zijn eigen voorkennis, opvattingen etc. Op dit laatste niveau kent een lezer aan een tekst betekenis toe en legt hij onuitgesproken verbanden.

Een cruciale - en voor eye-tracking bijzonder relevante - aanname in de constructie-integratietheorie van Kintsch is dat lezers de inhoudelijke structuur van een tekst herkennen en reconstrueren en van meet af aan hoofd- en bijzaken onderscheiden. Volgens Kintsch kennen aandachtige lezers in hun zoektocht naar samenhang vrijwel automatisch aan tekstgedeelten met een hoog topisch gehalte extra belang toe (Kintsch, 1998, m.n. p. 174). Meerdere onderzoeken hebben aangetoond dat geoefende lezers aan kernzinnen meer leestijd besteden dan aan detailzinnen, zelfs al bij de eerste lezing ervan (Hyönä & Niemi, 1990; Kintsch, 1998; Lorch et al., 1985; Mross, 1989; Pressley & Afflerbach, 1995; Yeari et al., 2015). Selectief oog hebben voor de inhoudsstructuur van de tekst is dan ook een prominente strategie van de expert-lezer (bijv. Duke & Pearson, 2002; McNamara, 2007; Perfetti et al., 2005).

1.3 Onderzoekshypothesen

Doel van deze studie is enerzijds meer zicht te krijgen op het tekstbegrip van vwo bovenbouw-leerlingen, en anderzijds op het verband tussen het vooraf-lezen en het vragen beantwoorden bij de dominerende taak in het Nederlandse tekstbegripsonderwijs: de tekst met vragen. Uitgangspunt zal zijn dat lezers, zoals vaak bij deze taak verondersteld, de tekst eerst aandachtig bestuderen en daarbij proberen een adequaat situatiemodel op te bouwen: ze zullen duidelijk aan kernzinnen meer leestijd schenken dan aan niet-kernzinnen. Daarnaast veronderstellen we een sterke samenhang tussen het lezen vooraf en het vragen beantwoorden.

We onderzoeken dit bij vwo 4-leerlingen, leerlingen die bij tamelijk complexe teksten onderscheid zouden moeten kunnen aanbrengen tussen hoofd- en bijzaken – dat stelt althans de Expertgroep Doorlopende Leerlijnen Taal (2008, 2009) over lezers op niveau 3F (vwo 4-leerlingen).

Voor het *vooraf-lezen* verwachten we allereerst dat vwo bovenbouw-leerlingen over het algemeen adequate tekstbases en situatiemodellen zullen opbouwen: ze zullen dus meer verwerkingstijd (leestijd) wijden aan kernzinnen dan aan niet-kernzinnen.

Ten aanzien van het *beantwoorden* vertrekken we vanuit de hypothese dat de participanten tijdens vooraf-lezen minimaal een adequate tekstbasis hebben opgebouwd en dus weten waar het antwoord op een vraag te vinden is: ze zullen aanzienlijk meer leestijd besteden aan de tekstdelen die voor het antwoord relevant zijn dan aan de tekstdelen die hiervoor irrelevant zijn. Ten tweede: de participanten hebben tijdens het vooraf-lezen een adequaat situatiemodel opgebouwd en zullen dus snel en adequaat antwoorden.

Voor de *relatie tussen het vooraf-lezen en het beantwoorden* formuleren we twee verwachtingen. Ten eerste: lezers die tijdens het vooraf-lezen langer naar kernzinnen keken en korter naar niet-kernzinnen, zullen effectiever antwoorden: ze zullen langer kijken naar de tekstdelen die voor het antwoord relevant zijn en korter naar tekstdelen die irrelevant zijn, vergeleken met lezers die kernzinnen niet positief onderscheiden van niet-kernzinnen (vgl. Rouet, 2006). Ten tweede: lezers die langer naar kernzinnen kijken en korter naar niet-kernzinnen, zullen vaker een goed antwoord op een vraag geven.

2 Methode

2.1 Participanten en experimentopzet

Twintig vwo 4-leerlingen van één grote collegiale middelbare school in het zuiden van Nederland werden aselekt gekozen. Drie van de gevraagde leerlingen besloten om uiteenlopende redenen niet te participeren. De zeventien deelnemende leerlingen hadden geen problemen met hun zicht. Zeven jongens namen deel, tien meisjes. De gemiddelde leeftijd was 16;3 jaar ($sd = 6$ maanden). Alle leerlingen

waren moedertaalsprekers Nederlands. Achteraf bleken de data van één participant geheel onbruikbaar, hoogstwaarschijnlijk vanwege het dragen van softlenzen, die de oogdetectie vaak belemmeren (vgl. Holmqvist & Andersson, 2017). Bij drie participanten werden data gedeeltelijk (doorgaans één tekstlezing) uitgesloten vanwege technische problemen.

Participanten ($N = 16$) bestudeerden op het computerscherm vier teksten, variërend in lengte van 240 tot 290 woorden. Na bestudering van de tekst volgden op het scherm telkens vier of vijf begripsvragen. Bij het maken van de begripsvragen bleef de tekst steeds naast de vraag raadpleegbaar zoals de praktijk is in het schoolvak Nederlands. Na elk tekstblok met vragen volgde een korte mogelijkheid tot pauze. De volgorde van de stimuli was voor alle deelnemers gelijk. Gedurende de hele *online* sessie volgde een eye-tracker de oogbewegingen van de deelnemers. Na het eye-tracking-experiment vulden de leerlingen een vragenlijst in, waarin naar enkele algemene gegevens werd gevraagd maar ook naar een inschatting van de motivatie en concentratie gedurende het experiment. Ten slotte volgde een kort retrospectief *stimulated recall*-interview: leerlingen werd hun antwoordgedrag tijdens de laatste set met vragen getoond en werd gevraagd hierop te reflecteren.

2.2 Stimuli

De inhoud en opzet van bovenbouwmethodes Nederlands en van het landelijk Centraal Schriftelijk Examen Nederlands vwo waren leidend in het samenstellen van de stimuli. We selecteerden uit landelijke dagbladen en tijdschriften vier korte opiniërende en uiteenzettende teksten, van maximaal 300 woorden lengte. De teksten liepen in moeilijkheidsgraad op van niveau 2F tot 4F. Bij elke tekst construeerden we vervolgens een acht- tot tiental vragen van wisselend niveau. Een screeningspanel, bestaande uit twee gepromoveerde docenten taalbeheersing en twee geroutineerde bovenbouwdocenten Nederlands, screende daarop al het materiaal uitvoerig en stelde het correctiemodel vast. Een papieren pretest-afname onder vier vwo 4-klassen volgde. Op basis van de leerlingantwoorden werden ten slotte per tekst vier of vijf items geselecteerd. Selectiecriteria waren in volgorde van gewicht: a) de kwaliteit van de vraagstelling (eenduidigheid, helderheid), b) de waarschijnlijkheid dat een vraag afwijkend oogbewegingengedrag genereerde tussen lezers met een goed en fout antwoord (baseerden leerlingen met een fout antwoord niet steeds hun antwoord op dezelfde tekstpassage als leerlingen met een goed antwoord?), en ten slotte c) statistische indicatoren (p-waarden en rit's).

De vragen werden zo geconstrueerd dat ze op zowel de tekstbasis als het situatiemodel betrekking hadden. Bij het merendeel van de vragen werd daarom in de stamvraag een voor de vraag relevant citaat uit de brontekst opgenomen, zonder vermelding van alinea- of regelnummer. Zo werd een beroep gedaan op de door de lezer opgestelde tekstbasis. Bij het construeren van de vragen werden *elaboration questions* en *bridging inference questions* geformuleerd, beide in open vraagvorm. *Elaboration questions* vragen om een tekstpassage te verbinden met de achtergrondkennis van een lezer; *bridging inference questions* vereisen dat een lezer twee

of meer tekstpassages inhoudelijk verbindt. Deze twee vraagtypen hebben op het situatiemodel betrekking (Cerdán et al., 2009; Kintsch, 1998; Land, 2009; O'Reilly & McNamara, 2007).

Door deze tweeledige inrichting van het item kon bij het beantwoorden worden verwacht dat er uiteenlopende oogbewegingspatronen optraden: leerlingen met een redelijke tot goede tekstbasis zullen het citaat uit de stamvraag direct in de tekst weten te lokaliseren, terwijl leerlingen met een matige tekstbasis de passage moeilijker of niet vinden. Leerlingen met een redelijke tekstbasis én een adequaat situatiemodel zullen hun aandacht geheel of grotendeels schenken aan de voor beantwoording noodzakelijke passages of zullen direct een antwoord geven, zonder de tekst te raadplegen (vgl. Rouet, 2006). In Bijlage 1 zijn de sets teksten en vragen opgenomen die als stimuli zijn gebruikt.

Omdat we willen weten of deelnemers tijdens het vooraf-lezen meer aandacht schenken aan kernzinnen, werd aan het screeningspaneel verzocht de kernzinnen in de vier geselecteerde teksten aan te wijzen. Daartoe ontvingen ze een uitgebreide schriftelijke instructie vooraf. De overeenstemming tussen de vier leden van dit paneel hierin bleek hoog (Cronbachs $\alpha = .82$). Daarnaast gaf het paneel bij elke vraag aan welke tekstpassages noodzakelijk zijn om de vraag correct te beantwoorden. Hierin bleken de vier leden van het paneel zeer sterk te overeenstemmen (Cronbachs $\alpha = .96$).

2.3 Procedure

Bij de afname op locatie is een mobiele SMI-eye-tracker gebruikt (versie 3.6) met een samplefrequentie van 250 Hz; bij de oogbewegingenregistratie van de eye-tracker is het rechteroog genomen. Als parameters voor de eventdetectie zijn conform Nyström en Holmqvist (2010) een minimale fixatieduur van 40 milliseconden gekozen en een minimale saccadeduur van 15 milliseconden. De SMI-eye-tracker kent op basis van saccadedetectie events toe; deze saccades werden toegekend vanaf een snelheidsdrempelwaarde van $40^\circ/\text{sec}$, een courante waarde bij leesonderzoek (vgl. Holmqvist & Andersson, 2017).

De deelnemers ontvingen vooraf een uitgebreide toelichting op het experiment door de onderzoeker. Daarin werd de werking van de eye-tracker kort uiteengezet en werd benadrukt dat het een algemeen onderzoek naar de leesontwikkeling van vwo-leerlingen betrof. De onderzoeker beklemtoonde daarnaast zeer uitdrukkelijk dat de getoonde teksten terdege bestudeerd dienden te worden, alsof er een toets over zou volgen. Daarna volgde een kalibratie- en validatieprocedure die, in verband met de enigszins ongebruikelijk lange tijdsduur van het eye-tracking-experiment, bij elke tekst herhaald werd. Deelnemers bepaalden zelf het tempo waarin de stimuli gepresenteerd werden en konden naar eigen believen naar een volgend scherm schakelen door de spatiebalk in te drukken; terugkeren naar een vorig scherm was niet mogelijk. Er werd vooraf geen maximumduur aan het experiment gesteld; deelnemers konden zelf bepalen hoeveel tijd zij aan lezing en beantwoording besteedden.

De zeventien deelnemers besteedden gemiddeld 27.5 minuten ($sd = 4.7$ min.) aan het *online* deel van het experiment. Het hele experiment met introductie, pauze en afronding duurde doorgaans een zestigtal minuten. De participanten gaven in de controlevragenlijst aan dat ze zich doorgaans tijdens het experiment ruim voldoende geconcentreerd voelden (op een tienpuntsschaal: gem. = 7.0; $sd = 1.2$) en ruim voldoende gemotiveerd (gem. = 7.5; $sd = 1.1$).

2.4 Data-analyse

Lezen vooraf. Voorafgaand aan de analyse werd de kwaliteit van de oogbewegingsdata bekeken en werden alle fixaties handmatig gecontroleerd via het programma Fixation (Cozijn, 2006). De kwaliteit van de data bleek over het algemeen hoog.

Om te kunnen onderzoeken in hoeverre vwo 4-leerlingen onderscheid maken tussen kernzinnen en detailzinnen, werden de vier teksten opgedeeld in *areas of interest* (aoi's), waarbij zinnen afzonderlijk gecodeerd werden als kernzinnen en niet-kernzinnen. Daarop brachten we twee factoren in kaart die de leestijden van deze aoi's kunnen beïnvloeden. Ten eerste: het ligt voor de hand dat deelnemers langere aoi's langer zullen bekijken dan kortere aoi's. Daarom werd per aoi de hoeveelheid karakters (letters incl. interpunctie) berekend. Ten tweede: lezers zullen doorgaans de eerste zinnen van een alinea langer bekijken dan de vervolgzinnen (Rayner, 1998). Daarom codeerden we de kernzinnen die vooraan in een alinea stonden tot een afzonderlijke categorie. Deze beide factoren zullen in de data-analyse worden verdisconteerd.

Voor de vier teksten werden zo in totaal 51 aoi's vastgesteld, waarvan er 18 betrekking hadden op kernzinnen en 33 op niet-kernzinnen; een kernzin-aoi bestond daarbij doorgaans uit één enkele zin, terwijl een niet-kernzin-aoi's vaak meerdere zinnen in de tekst overspande. Van de 18 kernzin-aoi's stonden er zes voorop in de alinea. De gemiddelde lengte van een tekst-aoi was 137 karakters.

Voor het vooraf-lezen van de teksten zijn vijf statistische modellen gespecificeerd. In het eerste model wordt een algemeen gemiddelde leestijd, de variantie tussen leerlingen, de variantie tussen teksten en de residuele variantie geschat. In de volgende modellen worden effecten een voor een toegevoegd zodat we het verschil in passing van de modellen eenduidig kunnen interpreteren. De aoi's verschillen in lengte; daarom wordt in het tweede model het effect van de hoeveelheid karakters op de leestijd van een aoi geschat. (Analyses vooraf hebben geleerd dat de relatie tussen kijktijd en lengte van de aoi niet-lineair is, en om die reden is niet alleen het lineaire maar ook het kwadratische lengte-effect geschat.) In het derde model schatten we de gemiddelde leestijden voor kernzinnen en niet-kernzinnen afzonderlijk. Omdat lezers voorop geplaatste zinnen in een alinea meestal meer leestijd schenken, wordt in het vierde model het effect van deze alinea-openende kernzinnen afzonderlijk meegenomen. In het vijfde en laatste model ten slotte wordt toegestaan dat het lengte-effect varieert tussen kernzinnen en niet-kernzinnen.

Deze vijf modellen worden toegepast op de (natuurlijke logaritme van de) *first-pass dwell time* en de (natuurlijke logaritme van de) totale kijktijd per aoi, bekende

leesmaten in de eye-tracking-literatuur. De *first-pass dwell time* geeft de totaaltijd weer van alle fixaties binnen een gecodeerde regio vóóordat er achteruit of vooruit wordt gekeken naar een andere regio, en zou daarmee indicatief zijn voor de eerste verwerking van een zin. De totale kijktijd geeft de totaaltijd van alle fixaties binnen een gecodeerde regio weer tijdens het gehele leesproces, verrekent dus alle regressies naar die regio toe, en incorporeert daarmee eerder complexere cognitieve processen zoals het oplossen van begripsproblemen (Holmqvist & Andersson, 2017).

Antwoord. De door participanten gegeven antwoorden bij de vragen werden gecorrigeerd door een van de onderzoekers aan de hand van het door het screeningspaneel vastgestelde correctiemodel; bij twijfel werden antwoorden voorgelegd aan leden van het screeningspaneel.

Vragen beantwoorden. Om het oogbewegingengedrag tijdens het beantwoorden van de vragen in kaart te brengen, ontwierpen we een tweede serie van *areas of interest*. De vraag zelf werd allereerst telkens in één aoi ondergebracht. Het screeningspaneel was vooraf gevraagd aan te geven welke tekstelementen voor de beantwoording relevant waren. Op basis hiervan konden we per vraag nog twee soorten aoi's onderscheiden: a) tekst-aoi's die voor beantwoording van de vraag relevant zijn, b) tekst-aoi's die voor beantwoording van de vraag irrelevant zijn. Omdat de lengte van een aoi waarschijnlijk van invloed zal zijn op de leestijd, bepaalden we voor elke aoi de hoeveelheid karakters. Ten slotte werd voor alle aoi's afzonderlijk de totale kijktijd berekend. De *first-pass dwell time* is hier overigens niet bruikbaar: ze veronderstelt een lineaire leeswijze en het beantwoorden kenmerkt zich juist door scannend zoekgedrag en een interactie tussen tekst en vraag (Brunfaut & McCray, 2015; Rouet, 2006).

Voor het beantwoorden ontstonden bij een tekst minimaal zeven, maximaal elf tekst-aoi's van minimaal één, maximaal vier tekstregels. In totaal werden er 33 tekst-aoi's onderscheiden, waarvan er naargelang de vraag doorgaans één of twee relevant waren voor de beantwoording. De achttien vragen zelf werden ondergebracht in achttien aparte aoi's. De gemiddelde lengte van een aoi bij het beantwoorden was 198 karakters.

Voor het beantwoorden zijn zes modellen gespecificeerd. In het eerste model wordt uitgegaan van een algemeen gemiddelde voor een antwoord-aoi. In de volgende modellen zijn effecten één voor één toegevoegd. Allereerst moet bij het beantwoorden rekening worden gehouden met het feit dat aoi's in lengte variëren; daarom wordt in het tweede model het algemene effect van de aoi-lengte geschat. In het derde model voegen we het effect van het antwoord-aoi-type toe; we bekijken de gemiddelde leestijd voor een vraag, voor een antwoord-relevante aoi en voor een antwoord-irrelevante aoi afzonderlijk. De lengte kan bij deze drie aoi-typen natuurlijk een wisselende invloed hebben op de leestijd; in het vierde model wordt daarom toegestaan dat het lengte-effect varieert tussen de drie aoi-typen. In het vijfde model wordt het effect van een goed antwoord verdisconteerd: is er een sa-

menhang tussen het geven van een goed dan wel fout antwoord en de leestijden in het antwoordproces? In het zesde model ten slotte wordt bekeken of het geven van een goed dan wel fout antwoord de leestijden van de drie antwoord-aoi-typen afzonderlijk beïnvloedt.

Om meer zicht te krijgen op de kwaliteit van het antwoorden, gebruikten we aanvullend een retrospectief *stimulated recall*-interview. Participanten werd hun antwoordgedrag bij de laatste set met vier vragen getoond en verzocht hierop te reflecteren. Daardoor kon aanvullend meer inzicht worden verkregen in het rationale achter hun antwoordgedrag, die immers nauwelijks vanuit eye-tracking-data is te deduceren (Holmqvist & Andersson, 2017). Uit de toelichting die participanten tijdens het interview gaven op hun eigen antwoordgedrag werd aan de hand van een observatieschema vastgesteld in hoeverre een participant direct na lezing van een vraag blijk gaf van a) een adequate tekstbasis, en b) een adequaat situatiemodel. Daarbij werd gescoord aan de hand van een vierpuntsschaal: apert aanwezig, voldoende aanwezig, enigszins aanwezig, niet aanwezig. Bij twijfel werden antwoorden voorgelegd aan leden van het screeningspanel.

Relatie tussen vooraf-lezen en het antwoord. Om de relatie tussen het vooraf-lezen en het gegeven antwoord te onderzoeken, werd eerst het aantal goede antwoorden per set met vragen berekend. Omdat de hoeveelheid vragen per set varieerde, berekenden we daarna de proportie van goede antwoorden per set. Deze proportie werd vervolgens gerelateerd aan de leestijden van kernzinnen en niet-kernzinnen bij het vooraf-lezen.

Daartoe zijn drie modellen gespecificeerd. In het eerste model wordt weer uitgegaan van het model met de beste *fit* bij het vooraf-lezen. In het tweede model wordt het effect van de proportie goede antwoorden op de vragen geschat op de leestijden bij het vooraf-lezen. In het derde model wordt het mogelijke interactie-effect tussen de proportie goede antwoorden en de leestijden van kernzinnen en (niet-)kernzinnen verdisconteerd.

3 Resultaten

3.1 Vooraf-lezen

Om het vooraf-lezen te beschrijven, werden twee leesmaten afzonderlijk geanalyseerd: 1) *first-pass dwell time*: de totaal tijd van alle fixaties binnen een regio vóórdat er voor- of achteruit wordt gekeken naar een andere regio; en (2) de totale kijktijd in een bepaalde regio. Wanneer een participant tijdens het lezen kort de laatste tekstregio fixeerte om daarna de overgeslagen aoi's lineair te lezen, kregen de overgeslagen aoi's bij de eerste leesmaat automatisch de waarde 0. Dit trad in 15% van het aantal aoi's op. Deze regio's werden uitgesloten uit de analyse van de *first-pass dwell time*. Regio's met incidentele 0-waarden bij de totale kijktijd (vrijwel steeds overgeslagen teksttitels) werden eveneens uitgesloten.

Tabel 2.1 Fit en passingsvergelijking van vijf modellen voor (de logaritme van) de first-pass dwell time en de totale kijktijd (-2LL: -2log likelihood) bij het vooraf-lezen

Model	-2LL	Modelvergelijking			
		Modellen	$\Delta\chi^2$	Δdf	p
<i>First-pass dwell time</i>					
1 intercept	2318.95				
2 + aoi-lengte	2043.34	2 vs 1	275.61	2	<.001
3 + kernzin	2043.11	3 vs 2	.23	1	.63
4 + kernzin vooraan in alinea	2038.75	4 vs 3	4.36	1	.04
5 + kernzin * aoi-lengte	2035.56	5 vs 4	3.19	1	.07
<i>Totale kijktijd</i>					
1 intercept	2030.70				
2 + aoi-lengte	843.70	2 vs 1	1187	2	<.001
3 + kernzin	827.34	3 vs 2	16.36	1	<.001
4 + kernzin vooraan in alinea	814.91	4 vs 3	12.43	1	<.001
5 + kernzin * aoi-lengte	813.5	5 vs 4	1.41	1	.24

Uit Tabel 2.1 blijkt dat de passing bij beide leesmaten drastisch verbetert wanneer we de lengte van de aoi's verdisconteren (Model 2 vs Model 1: $\Delta\chi^2(2) \geq 275.61$; $p < .001$). Wanneer we kernzinnen en niet-kernzinnen onderscheiden, genereert dat bij *first-pass dwell time* geen verbetering (Model 3 vs Model 2: $\Delta\chi^2(1) = .23$; $p = .60$), maar wel bij de totale kijktijd (Model 3 vs Model 2: $\Delta\chi^2(1) = 16.36$; $p < .001$). Toevoeging van de vooropplaatsing van de kernzin levert eveneens een betere passing op (Model 4 vs Model 3: $\Delta\chi^2(1) \geq 4.36$; $p \leq .04$). Het lengte-effect, waarin enkel het lineaire effect van de hoeveelheid karakters is betrokken, verschilt echter niet tussen kernzinnen en niet-kernzinnen (Model 5 vs Model 4: $\Delta\chi^2(1) \leq 3.19$; $p \geq .07$).

Tabel 2.2 geeft een schatting van de leestijden in *first-pass dwell time* en totale kijktijd voor kernzinnen en niet-kernzinnen, gebaseerd op model 4. Allereerst de *first-pass dwell time*: hier vinden we een duidelijk effect van de aoi-lengte, lineair ($\beta = .087$, $se = .006$) en kwadratisch ($\beta = -.002$, $se = .001$). Naarmate het aantal karakters toeneemt, kijken leerlingen duidelijk langer naar de zinnen in de tekst. Er is echter geen significant algemeen onderscheid aantoonbaar tussen kernzinnen en niet-kernzinnen ($\Delta\beta = -.047$, $se = .091$; $p = .60$). De kernzinnen die vooraan in de alinea staan, krijgen wel duidelijk meer aandacht dan niet-kernzinnen ($\Delta\beta = .277$, $se = .132$).

Dan de totale kijktijd: hier vinden we weer het te verwachten effect van aoi-lengte, zowel lineair ($\beta = .105$, $se = .004$) als kwadratisch ($\beta = -.003$, $se < .001$). We vinden een (nipt) significant onderscheid tussen kernzinnen en niet-kernzinnen: niet-kernzinnen worden lichtjes korter gelezen dan kernzinnen ($\Delta\beta = .072$, $se = .036$; $p = .05$). Staat een kernzin daarbij voorop in een alinea, dan wordt hij duidelijk langer bekeken dan een niet-kernzin ($\Delta\beta = .187$, $se = .053$).

Een vergelijking tussen de waarden bij beide leesmaten laat zien dat het effect van de aoi-lengte vrijwel gelijk blijft. Kernzinnen worden bij de *first-pass dwell time* niet langer bekeken dan niet-kernzinnen, maar wel lichtjes langer bij de totale kijktijd.

Tabel 2.2 Parameterschatting van de (logaritme van de) leestijd van zinnen (karakters) per leesmaat

Parameter	First-pass dwell time		Totale kijktijd	
	Regressiegewicht	(se)	Regressiegewicht	(se)
niet-kernzin	8.067	(.082)	8.572	(.053)
Δ kernzin	-.047	(.091)	.072	(.036)
Δ kernzin vooraan alinea	.277	(.132)	.187	(.053)
- lengte ^a	.087	(.006)	.105	(.002)
- lengte ²	-.002	(.001)	-.003	(.001)
<i>Varianties</i>				
S ² tekst	.003	(.001)	.061	(.003)
S ² leerling	.056	(.031)	.155	(.010)
S ² residu	.988	(.053)	.408	(.009)

^a lengte: gecentreerd rond gemiddelde (136.8 karakters) en gedeeld door 10

Kernzinnen die voorop staan in de alinea, worden daarentegen zowel in de *first-pass dwell time* als in de totale kijktijd duidelijk merkbaar langer bekeken.

Vergelijken we de gemiddelde (niet-)kernzinwaarden bij de beide leesmaten, dan lijken die van de totale kijktijd duidelijk hoger te liggen: na eerste lezing lijken leerlingen relatief vaak terug te kijken naar eerdere aoi's. Hierin dient nog verdisconteerd dat de *first-pass dwell time* de (geringe) duur van de saccades meeneemt en de totale kijktijd niet. Het aandeel terugkijken naar eerdere aoi's in de oogbewegingen ligt dus nog wat hoger.

Wordt er veel teruggekeken naar eerdere aoi's? De correlatie tussen beide leesmaten is hoog ($r = .61$; $p < .001$). Wanneer we deze correlatie atteneren voor onbetrouwbaarheid, dan blijken beide leesmaten niet significant van elkaar af te wijken ($r = 1.0$). Er is geen reden om *first-pass dwell time* en de totale kijktijd als twee verschillende leesprocessen te beschouwen: de hoeveelheid regressie naar eerdere aoi's tijdens het gehele proces is te verwaarlozen.

Ten slotte: de residuele variantie bij beide leesmaten is hoog, vooral bij *first-pass dwell time*. Daarvoor zijn meerdere verklaringen. Een deel van deze variantie ontstaat door de interactie tussen leerling en tekst en door een gedeelte error. Een ander belangrijk deel wordt waarschijnlijk veroorzaakt door het tamelijk onsystematische karakter van oogbewegingen, dat zich nog het sterkst voordoet bij de *first-pass dwell time*. Men denke aan 'standaard' regressies (Rayner, 1998): lezers fixeren kort een volgende zin, om dan vanuit technische leesproblemen even terug te keren naar de zojuist gelezen eerdere zin. In dat opzicht genereert elke zin ander, individueel bepaald leesgedrag. Het gedrag van leerlingen daarentegen is betrekkelijk stabiel te noemen, terwijl de verschillen ten gevolge van teksten relatief gering zijn.

3.2 Vragen beantwoorden

Eye-track-resultaten. Voor de beschrijving van het beantwoorden werd maar één leesmaat in de analyse betrokken: de totale kijktijd. Uit Tabel 2.3 blijkt dat de pas-

Tabel 2.3 Fit en passingsvergelijking van zes modellen voor de (logaritme van de) totale kijktijd (-2LL: -2log likelihood) bij het beantwoorden van de vragen

Model	-2LL	Modelvergelijking			
		Modellen	$\Delta\chi^2$	Δdf	<i>p</i>
1 intercept	7746.36				
2 + aoi-lengte	7734.79	2 vs 1	11.57	2	.003
3 + aoi-type	6553.46	3 vs 2	1181.33	2	<.001
4 + aoi-lengte * aoi-type	6548.97	4 vs 3	4.49	2	.11
5 + correct antwoord	6545.26	5 vs 4	3.71	1	.05
6 + correct antwoord * aoi-type	6529.77	6 vs 5	15.49	2	<.001

sing van een model waarin de lengte van de antwoord-aoi's wordt meegenomen, verschilt van het intercept (Model 2 vs Model 1: $\Delta\chi^2(2) = 11.57$; $p = .003$). Toevoeging van de drie antwoord-aoi-typen levert een duidelijk betere passing op (Model 3 vs Model 2: $\Delta\chi^2(2) = 1181.33$; $p < .001$), maar we vinden geen verbetering wanneer we het effect van de lengte laten variëren tussen de drie aoi-typen (Model 4 vs Model 3: $\Delta\chi^2(2) = 4.49$; $p = .11$). Een model dat de invloed van het geven van een goed antwoord op de leestijden in het antwoord meeneemt, heeft wel een betere passing (Model 5 vs Model 4: $\Delta\chi^2(1) = 3.71$; $p = .05$), evenals een model waarin we het effect van een goed antwoord laten verschillen tussen de drie aoi-typen (Model 6 vs Model 5: $\Delta\chi^2(2) = 15.49$; $p < .001$).

Tabel 2.4 Parameterschatting voor de (logaritme van de) totale kijktijd van antwoord-aoi's

Parameter	Coëfficiënt	(se)	<i>p</i>
vraag	9.776	(.124)	<.001
antwoord-relevante aoi	8.261	(.139)	<.001
antwoord-irrelevante aoi	7.135	(.092)	<.001
- aoi-lengte ^a	.062	(.012)	<.001
- aoi-lengte ²	-.001	(.000)	<.001
<i>Correct antwoord</i>			
Δ vraag * correct antwoord	-.296	(.127)	.02
Δ antwoord-relevante aoi * correct antwoord	.206	(.101)	.04
Δ antwoord-irrelevante aoi * correct antwoord	-.232	(.078)	<.01
<i>Varianties</i>			
S ² tekst	.027	(.013)	
S ² leerling	.026	(.012)	
S ² residu	1.024	(.031)	

^a lengte: gecentreerd rond gemiddelde (189 karakters) en gedeeld door 10

Tabel 2.4 geeft een schatting van de leestijden in totale kijktijd, gebaseerd op model 6, waarbij de toegevoegde parameter van model 4 (het effect van lengte verschilt tussen aoi-typen) is weggelaten. Voor het beantwoorden vinden we een duidelijk lengte-effect, in zowel lineair als kwadratisch opzicht (resp.: $\Delta\beta = .062$, $se = .012$; $\Delta\beta = -.001$, $se < .001$). Het effect van lengte wisselt echter niet tussen antwoord-relevante en -irrelevante aoi's. Uit de schattingen voor de drie antwoord-aoi-typen blijkt dat leerlingen tijdens het beantwoorden gemiddeld het langst naar de vraag kijken. En over het algemeen kijken leerlingen gemiddeld langer naar relevante antwoord-aoi's dan naar niet-relevante aoi's.

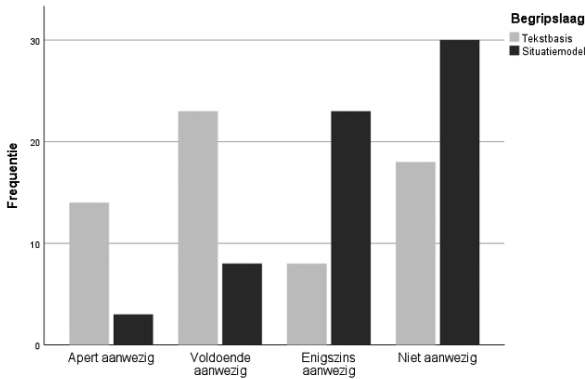
Er treedt een significant interactie-effect op tussen de antwoord-aoi-typen en het geven van een goed antwoord. Leerlingen die een goed antwoord geven, kijken relatief korter naar de vraag dan leerlingen met een fout antwoord ($\Delta\beta = -.296$, $se = .127$; $p = .02$). Leerlingen met een goed antwoord kijken bovendien langer naar antwoord-relevante aoi's in de tekst ($\Delta\beta = .206$, $se = .101$; $p = .04$) en korter naar antwoord-irrelevante aoi's ($\Delta\beta = -.232$, $se = .078$; $p < .01$) dan leerlingen met een fout antwoord.

Aanvullend hebben we een extra model opgesteld waarbij op leerlingniveau de covarianties geschat zijn tussen de vijf factoren bij het vooraf-lezen en beantwoorden: aan de ene kant kernzinnen en niet-kernzinnen bij het vooraf-lezen, aan de andere kant de drie aoi-typen bij het beantwoorden.

Dan blijkt het volgende: als leerlingen bij het vooraf-lezen kernzinnen snel of langzaam lezen, lezen ze ook de niet-kernzinnen snel of langzaam ($r = 1.0$). Als leerlingen bij het beantwoorden een vraag snel/langzaam lezen, lezen ze een antwoord-relevante of een antwoord-irrelevante aoi ook snel/langzaam ($r \geq .69$). En als leerlingen bij het vooraf-lezen de kernzinnen of niet-kernzinnen snel/langzaam lezen, lezen ze bij het antwoorden ook de vraag snel/langzaam ($r \geq .79$). We vinden dus steeds relatief hoge correlaties tussen de vijf factoren wanneer de leessnelheid van de participanten een verklaring vormt.

Is er ook een meer inhoudelijke samenhang aantoonbaar tussen vooraf-lezen en beantwoorden? Als een leerling bij het vooraf-lezen een kernzin of een niet-kernzin korter/langer leest, dan hangt dit niet significant samen met het korter dan wel langer lezen van een antwoord-relevante of -irrelevante aoi bij het beantwoorden ($r \leq .51$; $p \geq .12$). Anders gezegd: leerlingen die vooraf kernzinnen extra aandacht geven, hebben tijdens het beantwoorden niet significant meer oog voor antwoord-relevante tekstdelen. En leerlingen die vooraf niet-kernzinnen langer aandacht geven, hebben niet significant meer oog voor antwoord-irrelevante tekstdelen. Een meer inhoudelijke samenhang tussen lezen en beantwoorden is dus niet aantoonbaar.

Retrospectief interview. In het interview zagen leerlingen hun antwoordgedrag terug bij de vier vraagitems uit de laatste tekst. De aanwezige onderzoeker vroeg hun aan te geven in hoeverre ze onmiddellijk na vraaglezing a) wisten waar in de tekst het antwoord te vinden was (tekstbasis), b) een antwoord in gedachten hadden (situatiemodel). Over het algemeen leverden de uitspraken van leerlingen tijdens het retrospectieve *stimulated recall*-interview geen interpretatieproblemen op. Zes re-



Figuur 2.1 Frequentie van de begripsscores voor tekstbasis en situatiemodel

flecties bij een item werden voorgelegd aan leden van het screeningspanel. Bij één leerling bood de reflectie op de tekstbasis bij één item onvoldoende aanknopingspunten. Deze werd daarom uitgesloten.

Er is een duidelijk verschil waarneembaar in de verdeling van de tekstbasisscores en situatiemodelscores bij de laatste set met vragen ($\chi^2(3) = 24.62; p = .001$). Zie Figuur 2.1. Leerlingen gaven over het algemeen blijk van een redelijk aanwezige tekstbasis, waarbij ‘apert aanwezig’ en ‘voldoende aanwezig’ de meerderheid vormen. Bij de situatiemodelscores overheersen juist de indicatoren ‘enigszins aanwezig’ en ‘niet aanwezig’. Het onderscheid tussen beide begripsslagen bij ‘apert aanwezig’ is sterk. Bij tekstbasis scoorden leerlingen herhaaldelijk ‘apert aanwezig’: ze waren in staat erg precies een in de vraag vermeld citaat te lokaliseren. Bij het situatiemodel trad dit slechts in drie gevallen op: slechts in uitzonderlijke gevallen gaven leerlingen na vraaglezing direct antwoord of bleken ze een antwoord in hun hoofd te hebben.

3.3 Relatie tussen vooraf-lezen en antwoord

Voor het verband tussen het vooraf-lezen en het antwoord (het aantal goede antwoorden) zijn beide leesmaten geanalyseerd: *first-pass dwell time* en totale kijktijd. Uit Tabel 2.5 blijkt dat de passing van een model waarin de proportie goede antwoorden per set wordt meegenomen, bij beide leesmaten niet verschilt van het basismodel (Model 2 vs Model 1: $\Delta\chi^2(1) \leq .49; p \geq .48$). Toevoeging van het interactie-effect tussen (niet-)kernzinnen en de proportie goede antwoorden levert evenmin verbetering op (Model 3 vs Model 2: $\Delta\chi^2(1) \leq .55; p \geq .46$).

Er kan dus voor beide leesmaten geen verband aangetoond worden met het gegeven antwoord.

Tabel 2.5 Fit en passingsvergelijking van drie modellen voor de (logaritme van de) *first-pass dwell time* en totale kijktijd (-2LL: -2log likelihood) voor de relatie tussen vooraf-lezen en het antwoord

Model	-2LL	Modelvergelijking			
		Modellen	ΔX^2	Δdf	<i>p</i>
<i>First-pass dwell time</i>					
1 intercept	2038.75				
2 + proportie correct	2038.55	2 vs 1	.20	1	.65
3 + proportie correct * (niet-)kernzin	2038.38	3 vs 2	.17	1	.68
<i>Totale kijktijd</i>					
1 intercept	814.91				
2 + proportie correct	814.42	2 vs 1	.49	1	.48
3 + proportie correct * (niet-)kernzin	813.87	3 vs 2	.55	1	.46

4 Discussie en conclusie

In deze studie onderzochten we het vooraf-lezen en vragen beantwoorden bij teksten met vragen bij zestien vwo 4-leerlingen. We bespreken eerst het vooraf-lezen, vervolgens gaan we in op het beantwoorden en de relatie tussen het vooraf-lezen en het beantwoorden; we eindigen met enkele algemene slotobservaties.

4.1 Vooraf-lezen

De als eerste geformuleerde onderzoekshypothese luidde: vwo bovenbouw-leerlingen bouwen tijdens het vooraf-lezen over het algemeen een adequate tekstbasis en adequaat situatiemodel op, en ze zullen dus meer leestijd wijden aan kernzinnen dan aan niet-kernzinnen. Er is bij de totale kijktijd inderdaad sprake van meer leestijd voor kernzinnen dan niet-kernzinnen, en daarmee ondersteunt dit onderzoek de bevindingen uit eerdere studies dat geoefende lezers bij het lezen van teksten deze twee zinstypen onderscheiden (Hyönä & Niemi, 1990; Kintsch, 1998; Yeari et al., 2015). In vergelijking met deze eerdere studies valt het in dit onderzoek op dat de positieve kernzindiscriminatie alleen bij de totale kijktijd optreedt en niet in de *first-pass dwell time*. Dat laatste troffen Yeari et al. (2015) wel aan in hun onderzoekcondities.

Deze bevinding kan daarmee de uitspraak onderbouwen van de Expertgroep Doorlopende Leerlijnen Taal (2008, 2009) dat leerlingen op niveau 3F onderscheid maken tussen hoofd- en bijzaken: over het algemeen is er in deze studie sprake van het scheiden van hoofd- en bijzaak. Wel dienen we terughoudend te zijn met deze conclusie. Deze studie toont enkel aan dat dit effect zich over de groep van de zestien leerlingen als geheel voordeed. Bovendien troffen we dit onderscheid aanmerkelijk minder geprononceerd aan in de *first-pass*-maat.

Andere uitspraken van de Expertgroep Taal lijken we vanuit deze studie eerder te kunnen relativeren. Binnen de subdomeinen ‘begrijpen’ en ‘interpreteren’ van zakelijk lezen wordt in niveaubeschrijving 3F verondersteld dat leerlingen bij het

lezen van lastige teksten enkele niet-lineaire leesstrategieën inzetten, zoals lastige tekstdelen herlezen, opgestelde betekenissen herzien en de tekst scannen om alineafuncties te bepalen – bekende strategieën van expert-lezers (Duke & Pearson, 2002; Pressley & Afflerbach, 1995), strategieën ook die in bovenbouwmetho- des Nederlands nadruk krijgen. De twee leesmaten *first-pass dwell time* en totale kijktijd blijken echter in leestijd niet duidelijk van elkaar af te wijken. Daardoor is het onaannemelijk dat de onderzochte groep vwo 4-leerlingen deze strategische middelen bij het vooraf-lezen actief en veelvuldig heeft ingezet. Ook een ander in de marge ontwikkeld meetinstrument liet zien dat er nauwelijks sprake was van strategisch leesgedrag: leerlingen vertoonden bij het vooraf-lezen zelden ‘grote regressies’, ze sprongen, met andere woorden, zelden terug naar eerdere tekstdelen (Hyönä et al., 2002).¹ We kunnen wel stellen dat leerlingen vooral lineair lezen.

We moeten voorzichtig zijn hiermee harde conclusies te trekken over de kwaliteit van opgebouwde tekstbases en situatiemodellen bij het vooraf-lezen van de vwo 4-leerlingen: immers, vanuit de Kintschiaanse basisaanname van dit onderzoek – lezers maken onderscheid tussen kernzinnen en niet-kernzinnen – mogen we aannemen dat er een zekere tekstbasis en situatiemodel in het algemeen aanwezig zijn. Maar de overige bevindingen wijzen vooral op een tamelijk oppervlak- kig leesproces, zoals dat in enkele andere studies bij lezende vwo-leerlingen is aangetroffen (Stevens, 2018; Breukink, In voorbereiding). De vraag kan dan ook worden opgeworpen of het hier beschreven leesgedrag, zeker bij de moeilijkere teksten, volstaat om de teksten adequaat te doorgronden.

4.2 Vragen beantwoorden

Voor het beantwoorden vertrokken we in deze studie vanuit twee hypothesen. Allereerst: de leerlingen bouwen tijdens het vooraf-lezen ten minste een adequate tekst- basis op en besteden dus aanzienlijk meer tijd aan antwoord-relevante tekstdelen dan aan antwoord-irrelevante tekstdelen. Inderdaad, de deelnemers schenken bij de vragen meer aandacht aan antwoord-relevante dan aan antwoord-irrelevante tekstdelen. Tegelijk is de gemiddelde tijd die aan niet-relevante tekstdelen wordt besteed aanzienlijk: leerlingen kunnen dus eventueel de vereiste tekstbasis pas tij- dens het beantwoorden hebben opgebouwd.

Nadere duiding biedt hier het retrospectieve interview, dat inzicht geeft in het rationale achter het antwoordgedrag bij de laatste set met vragen. Direct na de vraaglezing blijkt er over het algemeen sprake van een redelijk aanwezige tekstba- sis: participanten geven bij de vier bevraagde items aan dat ze na vraaglezing door- gaans weten in welke alinea zich het antwoord bevindt. We lijken dus te mogen stellen dat leerlingen tijdens vooraf-lezen inderdaad een redelijke tekstbasis con- strueren.

1 Een regressie werd als een ‘grote regressie’ beschouwd wanneer er meer dan twee tekstregels op het scherm (> 100 karakters) werd teruggekeken. In 64% van de geobserveerde tekstlezingen trad er geen of slechts één grote regressie op; slechts drie tekstlezingen kenden vijf of meer grote regressies.

De tweede hypothese – dat leerlingen tijdens vooraf-lezen een adequaat situatiemodel hebben opgebouwd en daarmee snel en adequaat antwoorden – kunnen we vanuit de eye-track-data niet duidelijk beantwoorden. In het interview echter geven leerlingen bij de vier bevraagde vraagitems meestal aan dat ze het voor het antwoord noodzakelijke situatiemodel pas construeren tijdens het beantwoorden van de vraag. Dat versterkt de eerdere conclusie bij het vooraf-lezen dat leerlingen tamelijk oppervlakkig lezen.

We vonden verder in deze studie, zoals valt te verwachten, een duidelijke relatie tussen het beantwoorden en het uiteindelijke antwoord: leerlingen met een goed antwoord kijken beduidend langer naar relevante tekstdelen dan leerlingen met een onjuist antwoord; leerlingen met een goed antwoord kijken ook beduidend korter naar irrelevante tekstdelen. Ze kijken bovendien korter naar de vraag dan leerlingen met een onjuist antwoord. De verklaring lijkt eenvoudig: leerlingen die de vraag snel doorgronden, begrijpen beter waar het antwoord te vinden is en formuleren gemakkelijker een goed antwoord, terwijl leerlingen die veel moeite hebben met de vraag, vaker de vraag lezen, meer moeite hebben het antwoord in de tekst te vinden en ook minder snel een goed antwoord hebben.

4.3 *Relatie tussen vooraf-lezen en beantwoorden*

Voor de *relatie tussen vooraf-lezen en beantwoorden* formuleerden we op situatiemodelniveau twee verwachtingen. Ten eerste: lezers die tijdens het vooraf-lezen langer naar kernzinnen hebben gekeken dan naar niet-kernzinnen, zullen bij het beantwoorden van vragen langer kijken naar de voor beantwoording noodzakelijke tekstdelen, dan lezers die kernzinnen *niet* positief onderscheiden. We hebben deze hypothese niet kunnen bevestigen. Waar we in de analyse van de covarianties wel hoge correlaties binnen het vooraf-lezen aantreffen én binnen het beantwoorden, vonden we tussen het lezen en beantwoorden enkel een correlatie tussen het vooraf-lezen en het lezen van de vraag, en daarvoor vormt leessnelheid een voor de hand liggende verklaring. Maar voor het overige gedragen vooraf-lezen en beantwoorden zich hier als twee op zichzelf staande gehelen.

De tweede verwachting luidde: lezers die langer naar kernzinnen en/of korter naar niet-kernzinnen kijken, zullen vaker een goed antwoord op een vraag geven. Deze verwachting konden we evenmin bevestigen. Leerlingen met een goed antwoord kijken bij het vooraf-lezen over het algemeen niet korter of langer naar de tekst, noch kijken ze korter of langer naar kernzinnen en niet-kernzinnen.

We konden al met al op situatiemodelniveau geen relatie aantonen tussen het vooraf-lezen en het beantwoorden.

4.4 *Slotobservaties*

Enkele algemene voorbehouden bij deze studie zijn op hun plaats. Het betreft allereerst een oogbewegingsonderzoek: de aanname dat kijken naar een aoi een directe weerspiegeling is van de aandacht die een participant aan deze aoi schenkt,

staat nog altijd ter discussie (vgl. Holmqvist & Andersson, 2017); eye-tracking biedt daarnaast alleen zicht op het *hoe* en niet op het *waarom* achter lezen. Weliswaar vulden we deze lacune deels door een retrospectief interview in te zetten, dit had enkel betrekking op de laatste vier vragen bij een tekst op niveau 3F. Daarnaast betreft het een onderzoek naar het lezen en antwoorden op één Nederlandse school, bij slechts vier relatief korte teksten en een beperkt aantal vragen, onder zestien proefpersonen.

Overigens, vanwege de arbeidsintensiviteit van de onderzoeksmethode zijn de steekproeven bij vergelijkbare *mixed method*-eye-trackstudies niet veel groter (vgl. Bax, 2013; Brunfaut & McCray, 2015). De technische uitval van participanten is bovendien voor een eye-tracking-studie laag: vanwege de voor eye-tracking bijzondere lengte van het experiment kozen we voor een meermaals herhaalde kalibratie- en validatieprocedure die zeker aan de hoge kwaliteit van de data heeft bijgedragen. Deelnemers schatten ook hun motivatie en concentratie tijdens het experiment meestal in als ruim voldoende tot goed; incidentele opgaven van onvoldoende motivatie en/of concentratie lijken de data niet significant te hebben beïnvloed, zo bleek uit aanvullende analyses.

Inhoudelijk is het in deze studie dan allereerst opvallend dat we niet bij beide leesmaten meer leestijd voor kernzinnen vonden dan voor niet-kernzinnen, maar enkel bij de totale kijktijd. Daarvoor kunnen verschillende verklaringen zijn, zoals de beperktere leeservaring van de steekproef en vooral ook de heterogeniteit binnen de groep. Maar het kan ook samenhangen met de afwijkende experimentinrichting. Eerdere eye-track-studies naar kernzindetectie vertrokken vanuit *self-paced reading* (Hyönä & Niemi, 1990; Mross, 1989), óf vergeleken de leestijd van kernzinnen met die van perifere zinnen (Yeari et al., 2015). In deze studie is het natuurlijk lezen dichter benaderd dan bij *self-paced reading* en wordt de leestijd van kernzinnen vergeleken met de leestijd van de rest van de tekst. Dat we onder deze omstandigheden geen extra leestijd voor kernzinnen vinden bij het vooraf-lezen, kan komen doordat een lezer pas bij het opnieuw lezen prioriteit aan kernzinnen toekent.

Ten tweede: bij de tekst met vragen wordt doorgaans verondersteld dat het vooraf-lezen sterk het antwoord bepaalt. We konden op situatiemodel echter nauwelijks tot niet een relatie tussen het vooraf-lezen en het beantwoorden vaststellen. Wat we niet hebben kunnen aantonen, kan er natuurlijk weldegelijk zijn. Maar er tekent zich in de resultaten een duidelijk patroon af. We vonden enkel een voorzichtige ‘zwakke’ relatie tussen het vooraf-lezen en het beantwoorden op tekstbasisniveau. De participanten lazen vooraf vooral lineair, tamelijk oppervlakkig en bouwden waarschijnlijk pas tijdens het beantwoorden een meer dan rudimentair situatiemodel van de teksten op, ondanks de zeer nadrukkelijke instructie vooraf om de teksten terdege te bestuderen en het soms pittige karakter van de teksten. Het antwoordproces lijkt daarmee inhoudelijk vooral op zichzelf te staan.

Werkt de tekst met vragen dus oppervlakkig lezen in de hand, zoals vaak is beweerd (bijv. Hoogeveen & Bonset, 1998; SLO, 2016a, 2017)? Die kritiek lijkt vanuit deze studie terecht. Wij vonden in deze studie vrijwel geen diepere samenhang

tussen het vooraf-lezen en het beantwoorden van de vraag. En dat is misschien ook niet zo verwonderlijk: leerlingen weten bij deze taak nu eenmaal niet wat ze na het lezen zal worden gevraagd. In internationale leestaken en -toetsen wordt voorafgaande aan het lezen vaak een leesdoel en specifieke taakcontext gegeven, maar in methodes en toetsen in het schoolvak Nederlands ontbreekt deze meestal. Dat het beantwoorden zelf centraal komt te staan en niet het lezen, lijkt een vanzelfsprekend gevolg van de huidige inrichting van de tekst met vragen in het schoolvak Nederlands. De suggestie om voorafgaande aan leestoetsen steeds een context en doelstelling te formuleren lijkt daarom zinnig (vgl. Nederlands Nu! & Sectiebestuur Nederlands Levende Talen, 2018).

Bij eerdere studies bleek de tekst met vragen al een weinig valide toetsvorm om tekstbegrip op situatiemodelniveau te meten (Kamalski, 2007; Land, 2009). Vanuit deze studie is de vraag of leerlingen met deze taak wel leren lezen met dieper tekstbegrip: de vragen lijken de meeste aandacht te krijgen. De vraag is dus ook of deze taak wel zo geschikt is om tekstbegrip te trainen en of de huidige dominantie van de tekst met vragen in het schoolvak Nederlands als leestaak wel zo wenselijk is. Vervolgonderzoek zou zich moeten richten op de vraag of afwijkende inrichtingen van de tekst met vragen of alternatieve leestaken zoals de sorteertaak (Kamalski, 2007) geschikter zijn om valide tekstbegrip te oefenen en te trainen.

HOOFDSTUK 3

Teksten lezen, vragen beantwoorden,
geobserveerd bij vwo 2-, vwo 4- en
vwo 6-leerlingen

*Een cross-sectioneel oogbewegingsonderzoek**

* Dit hoofdstuk is in een eerdere vorm verschenen als: Rooijackers, P., Silfhout, G. van, Schuurs, U., Mulders, I., & Bergh, H. van den. (2020). Lezen en antwoorden bij teksten met vragen; een cross-sectionele eye-trackstudie onder 52 vwo-leerlingen. *Pedagogische Studiën*, 97(4), 281-308.

Samenvatting

In Nederland traint men vaak tekstbegrip met een tekst met vragen. Daarbij bestuderen leerlingen vooraf een tekst en beantwoorden vervolgens vragen, waarbij de tekst raadpleegbaar blijft. Bij deze taak werden verschillen in leesgedrag onderzocht tussen vwo 2- ($n = 17$), vwo 4- ($n = 16$) en vwo 6-leerlingen ($n = 19$).

Via oogbewegingenregistratie en retrospectieve interviews werd onderzocht in hoeverre jongere en oudere vwo-leerlingen ($N = 52$) in lees- en antwoordgedrag verschillen en hoe het vooraf-lezen van een tekst in deze jaarlagen de erop volgende vraagbeantwoording beïnvloedt.

De resultaten: bij de tekstbestudering vooraf doorbraken deelnemers slechts zelden het lineaire leespatroon; ze schonken aan kernzinnen nipt meer leestijd. Zesdeklassers en vierdeklassers lezen beduidend sneller dan tweedeklassers, maar zesdeklassers niet sneller dan vierdeklassers. Bij de vraagbeantwoording lokaliseerden zesdeklassers beter het antwoord in de tekst en antwoordden vaker correct dan tweedeklassers en (in mindere mate) vierdeklassers. In interviews gaven deelnemers aan bij vragen doorgaans te weten waar het antwoord in de tekst stond, maar nog geen (concept) antwoord paraat te hebben.

Deelnemers vertoonden ogenschijnlijk sterk pragmatisch leesgedrag. De vraag wordt opgeworpen in hoeverre deze taak oppervlakkig lezen in de hand werkt en of andere taken niet beter zelfstandig 'diep lezen' kunnen bewerkstelligen.

1 Inleiding

1.1 Aanleiding

Er is de afgelopen twintig jaar relatief veel onderzoek verricht naar het tekstbegrip van Nederlandse scholieren. Er zijn echter nog heel wat lacunes in onze kennis. Zo is het tekstbegrip van vmbo-scholieren weliswaar relatief scherp in kaart gebracht (bijv. De Milliano, 2013; Land, 2009; Schram, 2007), maar longitudinaal onderzoek naar hun ontwikkeling in tekstbegrip ontbreekt goeddeels. En er zijn slechts enkele deelstudies naar het tekstbegrip van vwo'ers (bijv. Van Silfhout, 2014). Er is ook internationaal leesonderzoek zoals PISA (Gubbels et al., 2019), dat 15-jarige vmbo- en havo- én vwo-leerlingen qua tekstbegrip internationaal op rang ordent. Maar voor vwo'ers ontbreken longitudinale of zelfs cross-sectionele studies naar hun ontwikkeling in tekstbegrip. Bij gebrek aan empirisch onderzoek naar tekstbegrip van vwo-leerlingen noemden de voorzitters van de Expertgroep Doorlopende Leerlijnen Taal (2008, 2009), het belangrijkste document over tekstbegrip in het Nederlandse secundair onderwijs, het gestelde eindniveau voor vwo 6 zelfs een "educated guess" (Robben, 2008).

Dit gebrek aan onderzoek naar de leesvaardigheid van vwo'ers is opvallend. Er zijn duidelijke zorgen over het tekstbegrip van instromende leerlingen op de Nederlandse universiteiten (Bonset, 2010; Raad voor de Nederlandse Taal en Letteren, 2015). Recent signaleerde het PISA-rapport over alle opleidingsniveaus in het voortgezet onderwijs een daling in tekstbegrip (Gubbels et al., 2019). En in het bijzonder tekstbegrip is een belangrijke voorwaarde voor het succes op vervolgstudies in het hbo en wo (bijv. Bogaert et al., 2008).

Al is er weinig onderzoek verricht naar het tekstbegrip van vwo-scholieren, de gemiddelde leesles Nederlands op het vwo laat zich redelijk scherp uittekenen. Uit vrij recent descriptief onderzoek naar de praktijk van de leeslessen Nederlands in onder meer vwo-klassen (Linthorst & De Glopper, 2015), een verkenning van de inrichting van de schoolexamens Nederlands lezen (Meestringa & Ravesloot, 2012) en de drie dominante huidige vwo-methodes Nederlands (Van der Leeuw et al., 2017) rijst een vrij eenduidig beeld op. In de gemiddelde vwo-leesles wordt weinig tot geen aandacht besteed aan technische leesvaardigheid of aan hardop lezen; in belangrijke mate domineert expliciete strategie-instructie de les. En mede onder invloed van het landelijk Centraal Examen Nederlands vwo speelt het individueel schrifte-

lijk beantwoorden van vragen bij een tekst een cruciale rol in de klas (vgl. Canton et al., 2013; Meestringa & Ravesloot, 2012; Pronk-van Eunen & De Vos, 2018; etc.).

Deze taak, de tekst met vragen, bepaalt in sterke mate wat momenteel bekend is over het tekstbegrip van vwo'ers – en dat is niet zonder problemen. Weliswaar kent de tekst met vragen een lange traditie in het schoolvak Nederlands (Wesdorp, 1981), de aloude, veelgehoorde kritiek erop luidt dat een lezer erdoor niet beter wordt in het doorgronden van teksten, maar alleen beter in het beantwoorden van begripsvragen (Hoogeveen & Bonset, 1998). Die kritiek, die de laatste jaren herhaaldelijk de media bereikte (bijv. Welgraven & De Vos, 2019), lijkt niet onzinnig. Leerlingen ontvangen bij deze taak doorgaans de algemene instructie een tekst vooraf terdege te bestuderen, en daarna dienen ze de tekstafhankelijke vragen te beantwoorden, waarbij de tekst raadpleegbaar blijft. Van oudsher wordt daarbij verondersteld dat wie vooraf goed de tekst heeft gelezen, de tekstafhankelijke vragen ook goed zal beantwoorden (vgl. Drop, 1979). Maar die aanname lijkt op gespannen voet te staan met internationaal onderzoek waaruit blijkt dat een specifiek leesdoel het lezen in sterke mate kan bepalen (Andre, 1979; Graesser et al., 1994; Yeari et al., 2015; etc.); bij de tekst met vragen is voor leerlingen het doel vooraf vaak vaag en impliciet ('Bestudeer de tekst en maak dan de vragen'), en pas ná eerste lezing van de tekst, bij het doornemen van de vragen wordt het concrete doel van de opdracht duidelijk. De vraag is dan of scores bij een tekst met vragen iets zeggen over de kwaliteit van het vooraf-lezen; weerspiegelen ze niet vooral de kwaliteit van de beantwoording van de begripsvragen?

Er is dus alle reden om het lees- en antwoordgedrag van vwo-scholieren, specifiek bij deze taak, nader te bestuderen. Dat doen we in deze studie op relatief non-invasieve wijze via eye-tracking, en we onderzoeken dit cross-sectioneel in drie jaarlagen (vwo 2, 4 en 6) met een in complexiteit oplopende set stimuli. Omdat eye-tracking geen inzicht biedt in het waarom achter lees- en antwoordgedrag, zetten we ondersteunend 'retrospectieve *stimulated recall*-interviews' in.

1.2 *Tekstbegrip van vwo-leerlingen in kaart*

Vwo-leerlingen behoren gemiddeld tot de betere lezers in het voortgezet onderwijs (vgl. Expertgroep Doorlopende Leerlijnen Taal, 2008, 2009; Gubbels et al., 2019; Van Silfhout, 2014; etc.). Wie verschillen in tekstbegrip tussen oudere en jongere vwo-leerlingen wil aantonen, zou dus logischerwijs moeten vertrekken vanuit de kenmerken van expert-lezers. Hun leesgedrag is redelijk scherp in kaart gebracht. In de studie van Pressley en Afflerbach (1995) zijn 38 hardopdenkstudies geanalyseerd om de algemene kenmerken van een volwassen lezer op te kunnen stellen. Een goede lezer blijkt zich vooral te kenmerken door de inzet van tekstbegrip bevorderende leesstrategieën, zoals leesdoelen stellen en voorkennis activeren tijdens het lezen, kenmerken die inmiddels hun weg hebben gevonden in tal van overzichtsstudies over tekstbegrip (Duke et al., 2011; McNamara, 2007; Perfetti et al., 2005; etc.).

Tekstbegrip is een verschijnsel dat zich enkel indirect laat beschrijven. Om zicht te krijgen op de constructie van begrip bij lezers, gebruikten onderzoekers in de

jaren tachtig en negentig van de twintigste eeuw vooral hardopdenkprotocollen, (vgl. Pressley & Afflerbach, 1995). Een belangrijk nadeel van hardopdenkprotocollen is echter dat de verbalisatie tijdens het lezen het natuurlijke handelingsproces nogal verstoort (bijv. Gerjets et al., 2011). Vanaf de jaren negentig wordt onder andere daarom oogbewegingenregistratie steeds vaker in onderzoek ingezet om het lezen te registreren, al richten deze studies zich wel vaak op deelaspecten van het lezen (bijv. Cozijn, 2000; De Leeuw et al., 2016; Kraal, 2020; Van der Schoot et al., 2012; Van Silfhout, 2014). Eye-tracking registreert waarop en hoelang de lezer diens ogen fixeert. De methode gaat ervan uit dat verschillen in de duur van fixaties verschillen in cognitieve verwerkingstijd vertegenwoordigen. Bij ‘saccades’, oogverplaatsingen, wordt de aandacht verschoven (Duchowski, 2007; Holmqvist & Andersson, 2017). Eye-tracking verstoort het natuurlijke lezen minder dan hardopdenken. Eye-track-data bieden evenwel geen zicht op het waarom achter het waargenomen leesgedrag (Bax, 2013; Cozijn, 2000).

Hoe toont men met eye-tracking verschillen aan in tekstbegrip tussen groepen jongere en oudere vwo-leerlingen? Een eerste, indirecte indicator kan de leessnelheid van een lezer zijn. Een snelle lezer is doorgaans een vaardiger lezer, een verschijnsel dat zich nog versterkt voordoet bij een orthografisch tamelijk regelmatige taal als het Nederlands (bijv. Jackson & McClelland, 1979; Perfetti, 1985; Wimmer, 1993). Vaardige lezers beschikken over het algemeen over een rijke woordenschat en een ruime situationele kennis en herkennen daarmee doorgaans vrijwel geautomatiseerd woorden. Daardoor lezen ze vaak relatief snel, waarmee ze meer cognitieve capaciteit beschikbaar hebben voor het oplossen van begripsproblemen dan minder vaardige lezers (Perfetti & Stafura, 2014). Natuurlijk hangt leessnelheid niet één op één samen met tekstbegrip (bijv. Wallot et al., 2014), maar goede lezers lezen over het algemeen wel sneller dan minder goede lezers. Kuijpers et al. (2003) vonden bijvoorbeeld dat Nederlandse volwassenen sneller lezen dan adolescenten en dat vwo-leerlingen gemiddeld sneller lezen dan mavoleerlingen. In de studie van Van Silfhout (2014) lazen de vwo-leerlingen sneller dan de vmbo-leerlingen. Alleen al om ontwikkelingspsychologische redenen zullen de vwo 6-lezers dus naar verwachting gemiddeld sneller lezen dan de vwo 2-lezers.

Een tweede mogelijke benadering is directer aan tekstbegrip gerelateerd: het al dan niet onderscheiden van kernzinnen. Volgens de Constructie-Integratietheorie van Kintsch (1998), uitgangspunt van vele studies naar leesvaardigheid en leesprocessen (vgl. McNamara & Magliano, 2009a), is een tekst een netwerk van hiërarchisch samenhangende, betekenisvolle beweringen (proposities), waarin een lezer relaties bouwt tussen woorden en zinnen (een microstructuur) en relaties tussen alinea's (een macrostructuur). Dit constructieproces mondt uit in een ‘tekstbasis’, een vrij letterlijke mentale representatie van de tekst, en een ‘situatiemodel’, een mentaal model waarin de lezer de tekstuele wereld integreert in zijn eigen voorkennis, opvattingen, etc. Pas op dit niveau kent een lezer aan een tekst diepere betekenis toe.

Een cruciale – en voor eye-tracking bijzonder relevante – aanname in Kintsch' theorie is dat aandachtige lezers vrijwel automatisch de inhoudsstructuur van een tekst herkennen en reconstrueren, hoofd- en bijzaken meteen onderscheiden.

Meerdere onderzoeken hebben aangetoond dat volwassen lezers aan kernzinnen meer tijd (aandacht) besteden dan aan detailzinnen, zelfs al bij eerste lezing ervan (Hyönä & Niemi, 1990; Kintsch, 1998; Lorch et al., 1985; Pressley & Afflerbach, 1995; Yeari et al., 2015). Oog hebben voor de inhoudsstructuur van de tekst is dan ook een prominente strategie van de expert-lezer (bijv. Duke & Pearson, 2002; McNamara, 2007; Perfetti et al., 2005). De Expertgroep Taal merkt op dat leerlingen op niveau 2F (grofweg vwo 2) in een tekst van dit niveau hoofd- en bijzaken kunnen scheiden; leerlingen op niveau 4F (grofweg vwo 6) zouden dit onderscheid ook bij teksten op niveau 4F moeten kunnen aanbrengen (Expertgroep Doorlopende Leerlijnen Taal, 2008, 2009). Als we leerlingen confronteren met in complexiteit wisselende teksten, ligt het daarom voor de hand aan te nemen dat vwo 6-leerlingen dit herkennen van kernzinnen duidelijker zullen laten zien dan vwo 2-leerlingen, vooral bij de eerste lezing van zinnen.

Een derde manier om met eye-tracking verschillen in tekstbegrip te tonen is verbonden aan de inzet van leesstrategieën. Zoals gezegd, kenmerken expert-lezers zich door het adequate gebruik van leesstrategieën. Veel van deze strategieën – zoals ‘goede lezers vergelijken en integreren hun voorkennis met de uitspraken in de tekst’ en ‘goede lezers construeren, reviseren en bevragen de concepten die ze tijdens het lezen opstellen’ (Duke et al., 2011) – kenmerken zich door een herlezing van eerdere zinnen en tekstdelen (bijv. Hyönä & Niemi, 1990). Vaardigere lezers teksten zullen dus minder lineair, minder zin voor zin lezen dan minder vaardige lezers (vgl. ook Schotter et al., 2014). En omdat strategie-instructie in de vwo-leesles Nederlands vaak een prominente plaats inneemt (bijv. Linthorst & De Gloppe, 2015), valt te verwachten dat oudere vwo-leerlingen strategieën eerder breed in zullen zetten bij een gedifferentieerd scala aan teksten, dan jongere vwo-leerlingen. Met andere woorden: een vwo 6-lezer zal vaker regressies naar eerdere zinnen en tekstdelen vertonen dan een vwo 2-lezer.

De beantwoording van vragen, waarbij we eveneens verschillen willen laten zien, staat wat theorievorming betreft nog in de kinderschoenen. Het antwoordproces wordt vaak omschreven als een grillig, door velerhande cognitieve processen veroorzaakt complex (vgl. Kintsch, 1998; Rouet, 2006). Bestaande studies belichten vaak de relatief eenvoudig meetbare opzoekvaardigheid in teksten naar aanleiding van vragen (Cerdán et al., 2009; Rouet, 2006; Rouet et al., 2001; Vidal-Abarca et al., 2011). In een paar studies werd de cognitieve validiteit van leestoetsen onderzocht (Bax, 2013; Brunfaut & McCray, 2015). Typerend genoeg bleek van de talloze hypothesen die Brunfaut en McCray (2015) over het beantwoorden bij een leestoets opstelden, een enkele empirisch aantoonbaar: vaardige lezers besteden minder tijd aan het maken van een vraag dan de minder vaardige lezers. Ontwikkelingspsychologisch ligt het voor de hand aan te nemen dat een vwo 6-leerling op dezelfde vraag bij dezelfde tekst sneller en adequater zal antwoorden en vaker een goed antwoord zal formuleren dan een vwo 2-leerling.

Ten slotte: het is essentieel het vooraf-lezen en het vragen beantwoorden ook op hun samenhang te onderzoeken. In het schoolvak Nederlands gaat men doorgaans ervan uit dat als leerlingen de tekst vooraf goed lezen, ze vervolgens vaker vragen

goed beantwoorden (bijv. Drop, 1979). Wie vooraf een tekst heeft gelezen, lijkt inderdaad beter te presteren op de aansluitende tekstbegripsvragen dan wie direct met de beantwoording van de vragen begint (Cerdán et al., 2009; vgl. ook Salmerón et al., 2015): leerlingen construeren waarschijnlijk vooraf een coherent beeld van de tekst, waardoor ze bij de vragen tot een dieper begrip van de tekst kunnen komen. In hoeverre *goed* lezen het resultaat hierbij bepaalt, is bij ons weten nog nauwelijks onderzocht. In hoofdstuk 2 bleken vwo 4-leerlingen tijdens vooraf-lezen vooral een redelijk scherpe tekstbasis op te bouwen, maar pas tijdens de vragen tot een rijker geschakeerd, dieper begrip van de tekst te komen. Vanuit de vigerende didactiek in methodes Nederlands ligt het voor de hand aan te nemen dat naarmate vwo-leerlingen ouder worden, ze een tekst vooraf grondiger en adequater hebben leren bestuderen: ze zouden dan beter moeten scoren bij de vragen. De vraag is dus in hoeverre we hier verschillen tussen jaarlagen kunnen aantonen: vinden we bij de oudere leerlingen een sterkere relatie tussen vooraf-lezen en vragen beantwoorden dan bij de jongere leerlingen?

1.3 Onderzoekshypothesen

Het doel van deze studie is een (cross-sectionele) ontwikkeling aan te tonen bij vwo-leerlingen bij teksten met vragen, de gebruikelijke leestaak in het schoolvak Nederlands in het vo. Daarom onderzoeken we cross-sectioneel het vooraf-lezen van vier teksten en het beantwoorden van bijbehorende vragen van vwo 2-, vwo 4- en vwo 6-leerlingen. Deze vwo-jaarlagen zijn grofweg te relateren aan resp. niveau 2F, 3F en 4F van de Expertgroep Doorlopende Leerlijnen Taal (2008, 2009). Daarbij instrueren we, analoog aan de praktijk in het schoolvak Nederlands, leerlingen vooraf dat ze de tekst eerst grondig hebben bestudeerd voordat ze de vragen gaan beantwoorden. Onze basishypothesen zijn de volgende: we verwachten dat het vooraf-lezen en het vragen beantwoorden een sterke relatie op tekstbasis- en situatiemodelniveau onderhouden (A). We gaan ervan uit dat naarmate leerlingen ouder worden, ze adequatere tekstbases en situatiemodellen opbouwen (B).

Daarvan uitgaande, poneren we dan voor *het vooraf-lezen* (1) de volgende drie verwachtingen: (1a) oudere vwo-leerlingen zullen in het algemeen sneller lezen dan jongere vwo-leerlingen (vwo 6 < vwo 4 < vwo 2); (1b) oudere leerlingen maken nadrukkelijker onderscheid tussen hoofd- en bijzaken: ze besteden relatief meer aandacht (kijktijd) aan kernzinnen dan aan niet-kernzinnen; (1c) oudere leerlingen lezen meer strategisch: ze herlezen vaker eerdere zinnen en tekstdelen en lezen dus minder lineair.

Voor *het vragen beantwoorden* (2) verwachten we dat oudere vwo-leerlingen adequater antwoorden dan jongere leerlingen: (2a) oudere leerlingen hebben vaker een goed antwoord; (2b) ze geven gemiddeld sneller antwoord; (2c) ze kijken vaker en langer naar de voor het antwoord relevante tekstdelen; en (2d) ze kijken korter en minder vaak naar de voor het antwoord irrelevante tekstdelen.

Voor *de relatie tussen vooraf-lezen en beantwoorden* (3) verwachten we, deels bij gebrek aan vooronderzoek, het volgende: (3a) als leerlingen bij het vooraf-lezen

relatief veel tijd besteden aan kernzinnen, dan kijken ze bij het beantwoorden ook langer en vaker naar de voor het antwoord relevante tekstdelen, (3b) als leerlingen relatief veel tijd besteden aan kernzinnen, kijken ze ook korter en minder vaak naar de voor het antwoord irrelevante tekstdelen; (3c) als lezers kernzinnen positief onderscheiden, hebben ze vaker een goed antwoord op de vragen; (3d) bij oudere vwo-leerlingen zijn deze voorgaande drie verwachtingen duidelijker aantoonbaar dan bij jongere vwo-leerlingen.

Eye-tracking-data staan alleen indirect uitspraken toe over de door lezers geconstrueerde tekstbases en situatiemodellen. Vooral wat de relatie tussen vooraf-lezen en vragen beantwoorden betreft zijn eye-tracking-data al snel meerduidig: heeft een leerling bijvoorbeeld snel het antwoord op een vraag geformuleerd doordat hij de tekst vooraf goed heeft begrepen óf doordat hij snel het antwoord in de tekst vond? Om de relatie tussen vooraf-lezen en vragen beantwoorden scherper te kunnen beschrijven, gebruiken we naast eye-tracking daarom 'retrospectieve *stimulated recall*-interviews', een werkwijze ontleend aan redelijk recente Britse studies naar antwoordgedrag (Bax, 2013; Brunfaut & McCray, 2015). Leerlingen wordt hun antwoordgedrag tijdens een deel van het experiment getoond en wordt gevraagd te reflecteren op de mate waarin ze hun antwoord baseerden op de initiële lezing van de teksten.

2 Methode

2.1 Participanten en experimentopzet

De afname van het experiment werd om technische redenen beperkt tot één grote collegiale middelbare school in het zuiden van Nederland. Deze school kent in de vwo-afdeling een voor Nederland representatieve leesdidactiek (zoals getypeerd in bijv. Linthorst & De Gloppe, 2015): de leidende methode Nederlands wordt in alle jaarlagen gebruikt, en leerlingen krijgen als taak vooral teksten met vragen. De cijfers voor het landelijk Centraal Schriftelijk Examen Nederlands tekstbegrip cirkelden de laatste jaren rondom het landelijk gemiddelde.

Twintig vwo 2-, twintig vwo 4- en twintig vwo 6-leerlingen werden aselekt gekozen. Alle deelnemers zijn moedertaalsprekers Nederlands. Vier leerlingen besloten om persoonlijke redenen niet te participeren. Achteraf bleken de data van vier leerlingen geheel onbruikbaar vanwege detectieproblemen. Bij vijf leerlingen werden data deels uitgesloten vanwege detectie- of technische problemen; doorgaans betrof het hier de detectie van een enkele gelezen tekst. In Tabel 3.1 is de uitval van de beoogde participanten over de drie jaarlagen veraanschouwelijkt. In totaal deden 22 jongens en 30 meisjes mee aan het experiment, redelijk evenwichtig verspreid over de drie jaarlagen. De gemiddelde leeftijd van de vwo 2-leerlingen was 13;10 jaar ($sd = 5$ maanden), van de vwo 4-leerlingen 16;2 jaar ($sd = 6$ maanden) en van de vwo 6-leerlingen 17;10 jaar ($sd = 7$ maanden).

Tabel 3.1 *Uitval van (data van) de aangezochte twintig participanten per jaarlaag*

(partiële) uitval	vwo 2 (n = 17)	vwo 4 (n = 16)	vwo 6 (n = 19)	totaal (N = 52)
geen toestemming	1	3	0	4
data geheel onbruikbaar	2	1	1	4
data deels onbruikbaar	2	3	0	5

Het experiment was voor alle jaarlagen identiek in opzet en inhoud. Participanten ($N = 52$) bestudeerden op het computerscherm vier teksten, variërend in lengte van 240 tot 290 woorden. Na het vooraf-lezen van de tekst volgden telkens vier of vijf vragen. Bij het maken van de vragen bleef de tekst steeds naast de vraag raadpleegbaar, zoals in het schoolvak Nederlands gebruikelijk. Leerlingen gaven hardop antwoord. De aanwezige onderzoeker noteerde hun antwoorden. Na elk tekstblok met vragen volgde een mogelijkheid tot pauze. Gedurende de hele *online* sessie werden de oogbewegingen van de deelnemers met een eye-tracker gevolgd. Na afronding van het eye-track-experiment vulden de leerlingen een controlevragenlijst in, waarin naar algemene gegevens werd gevraagd maar ook naar een inschatting van hun motivatie en concentratie gedurende het experiment. Ten slotte volgde een retrospectief *stimulated recall*-interview, waarbij leerlingen reageerden op een *scanpath* van hun antwoordgedrag bij de laatste set stimuli.

2.2 Stimuli

De inhoud en opzet van de methodes Nederlands en van het landelijk Centraal Schriftelijk Examen Nederlands vwo waren leidraad in het samenstellen van de stimuli. Uit landelijke dagbladen of tijdschriften selecteerden we vier korte opiniërende of uiteenzettende teksten, in moeilijkheidsgraad wisselend van niveau 2F tot 4F. Per tekst construeerden we acht tot tien vragen, eveneens in moeilijkheid wisselend. Door teksten en vragen in moeilijkheid te spreiden, kon worden verwacht dat dezelfde stimuli voor alle deelnemers inzetbaar waren.

Een screeningspanel, bestaande uit twee gepromoveerde docenten taalbeheersing en twee geroutineerde bovenbouwdocenten Nederlands, screende al het materiaal uitvoerig en stelde het correctiemodel vast. Daarna volgde een papieren pretestafname onder telkens twee vwo 2-, vwo 4- en vwo 6-klassen, waarna per tekst vier of vijf items werden geselecteerd. Selectiecriteria waren in volgorde van gewicht: a) de kwaliteit van de vraagstelling (eenduidigheid, helderheid); b) de waarschijnlijkheid dat een vraag afwijkend oogbewegingengedrag genereerde tussen lezers met een goed en fout antwoord (baseerden leerlingen met een fout antwoord niet steeds hun antwoord op dezelfde tekstpassage als leerlingen met een goed antwoord?); en ten slotte c) statistische indicatoren (p -waarden en rit's).

Vragen werden zo geconstrueerd dat ze op zowel de tekstbasis als het situatiemodel betrekking hadden. Bij het merendeel van de vragen werd daarom in de stamvraag een voor de vraag relevant citaat uit de brontekst opgenomen, zonder vermelding van alinea- of regelnummer. Zo werd een beroep gedaan op de door de lezer op-

gestelde tekstbasis. Bij het construeren van de vragen werden *elaboration questions* en *bridging inference questions* geformuleerd, beide in open vraagvorm. *Elaboration questions* vragen om de verbinding van een tekstpassage met de achtergrondkennis van een lezer; *bridging inference questions* vereisen dat een lezer twee of meer tekstpassages inhoudelijk verbindt. Deze twee vraagtypen hebben vooral op het situatiemodel betrekking (Kintsch, 1998; Land, 2009; O'Reilly & McNamara, 2007).

Door deze tweeledige inrichting van het item kon bij het beantwoorden worden verwacht dat er uiteenlopende oogbewegingspatronen optraden. Leerlingen met een redelijke tot goede tekstbasis zullen het citaat uit de stamvraag direct in de tekst weten te lokaliseren, terwijl leerlingen met een matige tekstbasis meer moeite ermee zullen ervaren de passage te vinden. Leerlingen met een redelijke tekstbasis én een adequaat situatiemodel zullen hun aandacht geheel of grotendeels schenken aan de voor beantwoording noodzakelijke passages of direct een antwoord geven of zullen direct een antwoord geven, zonder de tekst verder te raadplegen (vgl. Rouet, 2006). In Bijlage 1 zijn de sets teksten en vragen opgenomen die als stimuli zijn gebruikt.

Omdat we wilden weten of lezers meer aandacht schenken aan kernzinnen, verzochten we het screeningspanel de kernzinnen in de vier teksten aan te wijzen. Daartoe ontvingen ze vooraf een uitgebreide schriftelijke instructie. De overeenstemming tussen de vier leden van dit panel hierbij bleek hoog (Cronbachs $\alpha = .82$). Daarnaast gaf het panel per vraag aan welke tekstpassages relevant zijn om de vraag correct te beantwoorden. Hierin bleken de vier leden van het panel zeer sterk overeen te stemmen (Cronbachs $\alpha = .96$).

2.3 Procedure

Bij de afname op locatie is een mobiele SMI-eye-tracker gebruikt (versie 3.6) met een samplefrequentie van 250 Hz. Als parameters voor de eventdetectie zijn conform Nyström en Holmqvist (2010) een minimale fixatieduur van 40 milliseconden gekozen en een minimale saccadeduur van 15 milliseconden. De SMI-eye-tracker kent op basis van saccadedetectie events toe. Deze saccades werden toegekend vanaf een tamelijk courante snelheidsdrempelwaarde van 40° per seconde (vgl. Holmqvist & Andersson, 2017).

De deelnemers ontvingen vooraf een uitgebreide toelichting op het experiment door de onderzoeker. Deze zette de werking van de eye-tracker kort uiteen en benadrukte dat het een algemeen onderzoek naar de leesontwikkeling van vwo-leerlingen betrof. Hij beklemtoonde daarnaast zeer uitdrukkelijk dat de getoonde teksten terdege bestudeerd dienden te worden, alsof er een toets over zou volgen. Daarna volgde een kalibratie- en validatieprocedure die omwille van de detectiescherpte bij elke tekst herhaald werd. Deelnemers bepaalden zelf het tempo waarin de stimuli gepresenteerd werden en schakelden naar eigen believen naar een volgend scherm. Terugkeren naar een vorig scherm was niet mogelijk. Alle deelnemers kregen dezelfde stimuli in dezelfde volgorde voorgelegd. Er werd vooraf geen maximumduur aan het experiment gesteld.

Tijdens het retrospectief *stimulated recall*-interview kregen leerlingen een *scan-path* van hun antwoordgedrag te zien waarop ze werden uitgenodigd te reageren. De aanwezige onderzoeker vroeg daarbij na eerste lezing van een vraag steeds in hoeverre leerlingen wisten waar het antwoord stond en of ze al een antwoord in hun hoofd hadden.

De 52 deelnemers spendeerden gemiddeld 29.7 minuten ($sd = 7.9$) aan het *online* deel van het experiment. Het hele experiment, met introductie, pauze, interview en afronding, duurde per leerling doorgaans een zestigtal minuten.

2.4 Data-analyse

Vooraf-lezen. Voorafgaand aan de analyse werd de kwaliteit van de oogbewegingsdata bekeken en werden alle fixaties handmatig gecontroleerd via het programma Fixation (Cozijn, 2006). De koppeling van fixaties aan bekeken regio's leverde doorgaans weinig problemen op. De kwaliteit van de data was over het algemeen hoog.

Om te kunnen onderzoeken in hoeverre vwo-leerlingen onderscheid maken tussen kernzinnen en detailzinnen, zijn de vier teksten opgedeeld in een sequentie van *area's of interest* (aoi's), waarbij we deze als kernzin-aoi of niet-kernzin-aoi codeerden. Daarop brachten we twee factoren in kaart die de leestijden van deze aoi's kunnen beïnvloeden. Ten eerste: het ligt voor de hand dat deelnemers langere aoi's langer zullen bekijken dan kortere aoi's. Daarom werd per aoi de hoeveelheid karakters (lettertekens incl. interpunctie) berekend. Ten tweede: lezers zullen doorgaans de eerste zinnen van een alinea langer bekijken dan de vervolgzinnen (Rayner, 1998). Daarom codeerden we de kernzinnen die vooraan in een alinea stonden als een afzonderlijke categorie. Deze beide factoren zijn in de data-analyse verdisconteerd.

Voor de vier teksten werden zo in totaal 51 aoi's vastgesteld, waarvan er achttien betrekking hadden op kernzinnen en 33 op niet-kernzinnen. Een kernzin-aoi bestond daarbij doorgaans uit één enkele zin, terwijl een niet-kernzin-aoi vaak meerdere zinnen in de tekst overspande. Van de achttien kernzin-aoi's stonden er zes voorop in de alinea. De gemiddelde lengte van een tekst-aoi was 137 karakters.

Voor de analyse wordt vooral gebruikgemaakt van multiniveau-modellen waarin observaties steeds genest zijn binnen leerlingen en binnen teksten.

Voor het vooraf-lezen van de teksten zijn acht modellen gespecificeerd. In het eerste model schatten we een algemeen gemiddelde leestijd van een aoi (intercept), een variantie tussen leerlingen, een variantie tussen teksten en een residuele variantie. In de volgende modellen voegen we effecten één voor één toe zodat we het verschil in passing van de modellen eenduidig kunnen interpreteren. De aoi's verschillen in lengte; daarom wordt in het tweede model het effect van de hoeveelheid karakters op de leestijd van een aoi geschat. Analyses vooraf hebben geleerd dat de relatie tussen kijktijd en lengte van de aoi niet-lineair is, en om die reden is niet alleen het lineaire maar ook het kwadratische lengte-effect geschat. In het derde model schatten we dan de gemiddelde leestijd voor kernzinnen en niet-kern-

zinnen. Omdat lezers doorgaans meer leestijd besteden aan de voorop geplaatste zinnen in een alinea, schatten we in het vierde model nog afzonderlijk het effect van deze alinea-openende kernzinnen. In het vijfde en zesde model wordt bekeken of het effect van de lengte van aoi's wisselt tussen kernzinnen en niet-kernzinnen. En in het zevende model wordt onderzocht of de leestijden van aoi's in het algemeen per jaarlaag variëren, waarna we met het laatste model nagaan of er tussen jaarlagen verschillen bestaan in het lezen van kernzinnen en niet-kernzinnen.

Deze acht modellen worden toegepast op de (natuurlijke logaritme van de) *first-pass dwell time* en de (natuurlijke logaritme van de) totale kijktijd per aoi. De *first-pass dwell time* geeft de totaaltijd weer van alle fixaties binnen een gecodeerde regio vóórdat er achteruit of vooruit wordt gekeken naar een andere regio en is daarmee indicatief voor de eerste verwerking van een zin. De totale kijktijd geeft de totaaltijd van alle fixaties binnen een gecodeerde regio weer tijdens het hele leesproces; deze kijktijd verrekenet dus alle regressies naar deze regio en incorporeert eerder complexere cognitieve processen zoals het oplossen van begripsproblemen (Holmqvist & Andersson, 2017).

Om te bekijken in hoeverre leerlingen terugkijken naar eerdere zinnen in de tekst, kijken we op basis van het best passende model aanvullend nog naar de relatie tussen *first-pass dwell time* en totale kijktijd: in hoeverre wijken beide maten van elkaar af? Daarnaast berekenen we het percentage regressietijd binnen de totale kijktijd: welk deel van de totale kijktijd van een aoi ontstond door herlezing, i.e. nadat een lezer vanuit een erop volgende aoi ernaar terugkeek?

Het antwoord. De door participanten gegeven antwoorden bij de vragen werden gecorrigeerd door een van de onderzoekers met het door het screeningspaneel vastgestelde correctiemodel; bij twijfel werden antwoorden voorgelegd aan leden van het screeningspaneel.

Antwoordproces: eye-tracking. Om het oogbewegingengedrag tijdens het beantwoorden van de vragen in kaart te brengen, ontwierpen we een tweede serie van *area's of interest*. De vraag zelf werd allereerst telkens in één aoi ondergebracht. Het screeningspaneel was vooraf gevraagd aan te geven welke tekstelementen voor de beantwoording relevant waren. Op basis hiervan konden per vraag nog twee soorten aoi's worden onderscheiden: 1) tekst-aoi's die voor beantwoording van de vraag relevant zijn, 2) tekst-aoi's die voor beantwoording van de vraag irrelevant zijn. Omdat de lengte van een aoi waarschijnlijk van invloed zal zijn op de leestijd, bepaalden we voor elke aoi de hoeveelheid karakters. Ten slotte werd voor alle aoi's afzonderlijk de totale kijktijd berekend¹.

Voor het antwoordproces werden de achttien vragen dus ondergebracht in achttien afzonderlijke aoi's. Daarnaast ontwierpen we bij een tekst minimaal zeven,

¹ De *first-pass dwell time* is hier niet bruikbaar: ze veronderstelt een lineaire leeswijze en het beantwoorden kenmerkt zich juist door scannend zoekgedrag en een interactie tussen tekst en vraag (Brunfaut & McCray, 2015; Rouet, 2006).

maximaal elf tekst-aoi's van minimaal een, maximaal vier tekstregels. In totaal werden er 33 tekst-aoi's onderscheiden, waarvan er naargelang de vraag doorgaans een of twee relevant waren voor het antwoord. De gemiddelde lengte van deze 51 aoi's samen was 198 karakters.

Voor het antwoordproces zijn tien modellen gespecificeerd. In het eerste model schatten we een algemene gemiddelde leestijd van een antwoord-aoi (intercept), waarna we in de volgende modellen effecten één voor één toevoegen. Allereerst moet rekening worden gehouden met het feit dat de antwoord-aoi's in lengte variëren; daarom wordt in het tweede model het lineaire effect van de aoi-lengte geschat. (Initiële analyses hebben geleerd dat hier enkel van een lineair lengte-effect sprake is, niet kwadratisch.) In het derde model voegen we het effect van het antwoord-aoi-type toe: een gemiddelde leestijd wordt geschat voor een vraag, voor een antwoord-relevante aoi en voor een antwoord-irrelevante aoi. De lengte kan bij deze drie aoi-typen natuurlijk de leestijd wisselend beïnvloeden; in het vierde model wordt daarom toegestaan dat het lengte-effect varieert tussen de drie aoi-typen.

Met de modellen 5 tot en met 7 onderzoeken we de relatie met jaarlaag. In model 5 wordt bekeken of de leestijd van een antwoord-aoi verschilt tussen jaarlagen. Model 6 staat toe dat de leestijden van de drie afzonderlijke antwoord-aoi-typen wisselen tussen jaarlagen. En in model 7 nemen we de interactie tussen jaarlaag en aoi-lengte mee.

Modellen 8 tot en met 10 bekijken de invloed die het geven van een goed antwoord op de leestijden had. In hoeverre het geven van een goed dan wel fout antwoord in het algemeen de leestijd beïnvloedt, wordt in het achtste model bekeken. Model 9 splitst dit uit voor de drie antwoord-aoi-typen afzonderlijk. In model 10 ten slotte bekijken we of het geven van een goed dan wel fout antwoord per jaarlaag een wisselend effect heeft.

In bovenstaande analyse blijft echter een belangrijke factor buiten beschouwing. Niet alleen de leestijd van de drie aoi-types kan wisselen tussen jaarlagen, maar ook het aantal bekeken aoi's: leerlingen zullen waarschijnlijk per vraag naar meerdere antwoord-relevante en antwoord-irrelevante aoi's hebben gekeken. Om dat te onderzoeken, stelden we aanvullend een multinomiaal model op, waarin we per aoi-type bekijken of de hoeveelheid bekeken antwoord-relevante en antwoord-irrelevante aoi's wisselt tussen jaarlagen, en tussen leerlingen met een goed en een fout antwoord. Daarin verdisconteren we wederom de aoi-lengte en schatten we de variantie tussen leerlingen, de variantie tussen teksten en de residuele variantie.

Antwoordproces: retrospectief interview. Om meer duidelijkheid te krijgen in het waarom achter het waargenomen antwoordgedrag, gebruiken we aanvullend een retrospectief *stimulated recall*-interview. We willen hier vooral weten in hoeverre leerlingen direct na het lezen van de vraag, op basis van de initiële tekstbestudering al een antwoord in hun hoofd hadden. Uit de reflectie van participanten op het eigen antwoorden werd na afname aan de hand van een van tevoren vastge-

steld observatieschema vastgesteld in hoeverre een deelnemer meteen na lezing van een vraag blijk gaf van a) een adequate tekstbasis, en b) een adequaat situatiemodel. Daarbij werd gescoord aan de hand van een vierpuntsschaal: apert aanwezig, voldoende aanwezig, enigszins aanwezig, niet aanwezig.²

De door participanten gegeven antwoorden werden gecodeerd door een van de onderzoekers aan de hand van dit observatieschema; bij twijfel werden antwoorden voorgelegd aan leden van het screeningspanel.

Relatie tussen vooraf-lezen en antwoord. Om de relatie tussen het vooraf-lezen en het antwoord te onderzoeken, werd eerst het aantal goede antwoorden per set met vragen berekend. Omdat het aantal vragen per set wisselde, berekenden we vervolgens de proportie van goede antwoorden per set, die daarna werd gerelateerd aan de leestijden van kernzinnen en niet-kernzinnen bij het vooraf-lezen.

Daartoe zijn vier modellen gespecificeerd. In het eerste model wordt uitgegaan van het model met de beste *fit* bij het vooraf-lezen. In het tweede model wordt het effect van de proportie goede antwoorden geschat op de leestijden bij het vooraf-lezen. In het derde model wordt het mogelijke interactie-effect tussen de proportie goede antwoorden en de leestijden van kernzinnen en (niet-)kernzinnen verdisconteerd. In het vierde model verkennen we of er in dit opzicht een verschil is waar te nemen tussen jaarlagen: hebben leerlingen uit een jaarlaag meer of minder leestijd aan (niet-)kernzinnen besteed dan leerlingen uit een andere jaarlaag als ze een goed of fout antwoord hadden?

3 Resultaten

3.1 Algemeen

Allereerst: de deelnemers gaven aan dat ze zich tijdens de experimentafname gemiddeld ruim voldoende tot goed geconcentreerd voelden (op een tienpuntschaal: $M = 7.5$, $sd = 1.3$) en goed gemotiveerd ($M = 7.9$, $sd = 1.0$).

² Als concretisaties voor de scores bij de *tekstbasis* gold hierbij resp.: 1) 'zeer duidelijk aanwezig': weet een antwoord-relevant of een in de vraag geciteerd tekstdeel direct met een eventuele initiële afwijking van één regel te lokaliseren; 2) 'voldoende aanwezig': weet in welke alinea een antwoord-relevant tekstdeel of een in de vraag geciteerd tekstdeel zich bevindt; 3) 'enigszins aanwezig': weet of een antwoord-relevant tekstdeel of een in de vraag geciteerd tekstdeel zich in begin, midden of eind van de tekst bevindt; 4) 'niet aanwezig': weet niet waar een antwoord-relevant tekstdeel of een in de vraag geciteerd tekstdeel zich in de tekst bevindt. Als concretisaties voor de scores bij het *situatiemodel* gold resp.: 1) 'zeer duidelijk aanwezig': formuleert direct het antwoord of geeft aan direct het antwoord te weten; 2) 'voldoende aanwezig': geeft aan een idee of begin van een antwoord in het hoofd te hebben; 3) 'enigszins aanwezig': geeft weliswaar aan het antwoord nog niet te weten, maar weet wel direct waar het antwoord in de tekst te vinden is - we gaan hier uit van het idee dat een adequate tekstbasis doorgaans een rudimentair situatiemodel met zich meebrengt (vgl. Kintsch, 1998); 4) 'niet aanwezig': heeft geen idee wat het antwoord is, noch hoe het antwoord gevonden kan worden.

3.2 Vooraf-lezen

Passingsvergelijking. Voor de beschrijving van het vooraf-lezen werden met behulp van multiniveau-modellen twee leesmaten geanalyseerd: (1) *first-pass dwell time*: de totaal-tijd van alle fixaties binnen een gecodeerde regio vóórdat er voor- of achteruit wordt gekeken naar een andere regio; en (2) de totale kijktijd in een bepaalde regio. Wanneer een participant tijdens lezing kort de laatste tekstregio fixeerte om daarna de overgeslagen aoi's lineair te lezen, kregen de overgeslagen aoi's bij de eerste leesmaat automatisch de waarde 0. Deze regio's met 0-waarden (5.1% van het aantal aoi's) werden uitgesloten uit de analyse van de *first-pass dwell time*, evenals de incidentele 0-waarden bij de totale kijktijd (1%, vrijwel steeds overgeslagen teksttitels).

Tabel 3.2 Fit en passingsvergelijking van acht modellen voor (de logaritme van) de *first-pass dwell time* en de totale kijktijd (-2LL: -2log likelihood) bij het vooraf-lezen

Model	-2LL	Modelvergelijking			
		Modellen	$\Delta\chi^2$	Δdf	<i>p</i>
<i>First-pass dwell time</i>					
1 intercept	7675.83				
2 + lengte aoi ^a + lengte aoi ²	6395.83	2 vs 1	1280.00	2	<.001
3 + kernzin	6390.24	3 vs 2	5.59	1	.02
4 + kernzin vooraan in alinea	6385.42	4 vs 3	4.82	1	.03
5 + kernzin * lengte aoi ^a	6380.90	5 vs 4	4.52	1	.03
6 + kernzin * lengte aoi ²	6379.73	6 vs 5	1.17	1	.28
7 + jaarlaag	6355.34	7 vs 6	24.39	2	<.001
8 + jaarlaag * kernzin	6353.94	8 vs 7	1.40	2	.50
<i>Totale kijktijd</i>					
1 intercept	6901.82				
2 + lengte aoi ^a + lengte aoi ²	2883.73	2 vs 1	4018.09	2	<.001
3 + kernzin	2854.52	3 vs 2	29.21	1	<.001
4 + kernzin vooraan in alinea	2821.70	4 vs 3	32.82	1	<.001
5 + kernzin * lengte aoi ^a	2816.51	5 vs 4	5.19	1	.02
6 + kernzin * lengte aoi ²	2815.13	6 vs 5	1.38	1	.24
7 + jaarlaag	2803.36	7 vs 6	11.77	2	<.01
8 + jaarlaag * kernzin	2801.81	8 vs 7	1.55	2	.46

^a lengte: gecentreerd rond gemiddelde (189 karakters) en gedeeld door 10

Uit Tabel 3.2 blijkt dat de passing van een model met één algemeen gemiddelde voor beide leesmaten sterk verbetert wanneer daarin ook de aoi-lengte wordt meegenomen (Model 2 vs Model 1: $\Delta\chi^2(2) \geq 1280$; $p \leq .001$). Toevoeging van het onderscheid kernzin/niet-kernzin levert eveneens bij beide een betere passing op (Model 3 vs Model 2: $\Delta\chi^2(1) \geq 5.59$; $p \leq .02$), evenals de toevoeging van voorop geplaatste kernzinnen (Model 4 vs Model 3: $\Delta\chi^2(1) \geq 4.82$; $p \leq .03$). Een model waarin de leestijd van kernzinnen of van niet-kernzinnen afhankelijk is van de lineaire aoi-lengte, levert eveneens een verbetering op (Model 5 vs Model 4: $\Delta\chi^2(1) \geq 4.52$; $p \leq .03$); van

een interactie tussen (niet-)kernzinnen en de kwadratische aoi-lengte is bij beide leesmaten evenwel geen sprake (Model 6 vs Model 5: $\Delta\chi^2(1) \leq 1.38$; $p \geq .24$). Toevoeging van de jaarlaag zorgt bij beide leesmaten voor een aanmerkelijk betere passing (Model 7 vs Model 6: $\Delta\chi^2(2) \geq 11.77$; $p \leq .001$), maar er blijkt geen interactie-effect tussen jaarlaag en (niet-)kernzinnen (Model 8 vs Model 7: $\Delta\chi^2(2) \leq 1.55$; $p \geq .46$). Hypothese 1b kan dus bij beide leesmaten niet worden aangetoond: oudere vwo-leerlingen kijken niet beduidend langer naar kernzinnen dan jongere vwo-leerlingen.

Tabel 3.3 Parameterschattingen van de (logaritme van de) leestijd van zinnen per leesmaat

Parameter	First-pass dwell time		Totale kijktijd	
	Regressiegewicht	(se)	Regressiegewicht	(se)
niet-kernzin (vwo 6)	8.286	(.067)	8.561	(.056)
Δ kernzin	.032	(.045)	.035	(.021)
Δ kernzin vooraan in alinea	.119	(.065)	.157	(.030)
- lengte aoi ^a	.010	(.003)	.109	(.001)
- lengte aoi ²	-.003	(.001)	-.003	(<.001)
- kernzin * lengte aoi	-.012	(.005)	-.006	(.002)
Jaarlaag				
Δ vwo 2	.386	(.081)	.253	(.073)
Δ vwo 4	-.031	(.083)	.047	(.074)
Varianties				
S ² tekst	.003	(.003)	.002	(.002)
S ² leerling	.042	(.012)	.044	(.009)
S ² residu	.787	(.023)	.169	(.005)

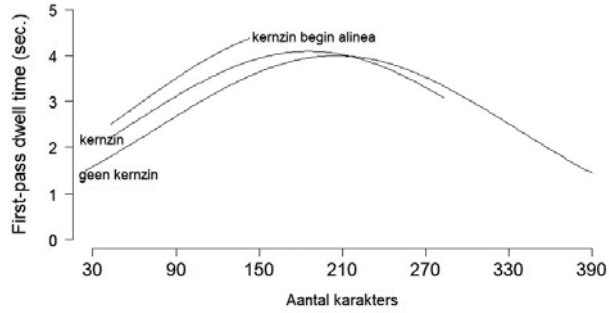
^a lengte: gecentreerd rond gemiddelde (189 karakters) en gedeeld door 10

Leestijden first-pass dwell time. Tabel 3.3 geeft een parameterschatting van de leestijden in *first-pass dwell time* en totale kijktijd voor kernzinnen en niet-kernzinnen, gebaseerd op model 7, waarbij niet-significante parameters buiten de analyse zijn gelaten (model 6: het interactie-effect tussen tussen (niet-)kernzinnen en kwadratische aoi-lengte).

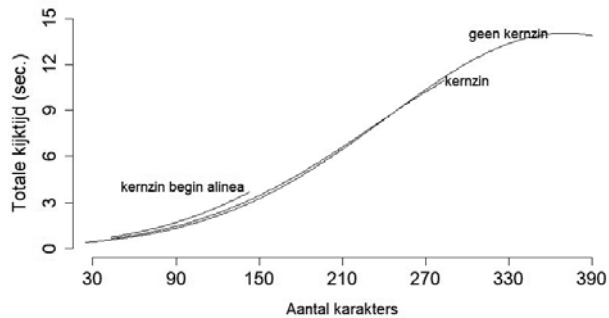
In Figuur 3.1 zijn de leestijden in de *first-pass*-maat voor de zesdeklassers weergegeven, gebaseerd op Tabel 3.3. Bij de *first-pass dwell time* hebben leerlingen (de zesdeklassers) gemiddeld $e^{8.286}$ (= 3968) milliseconden nodig om een niet-kernzin te lezen van gemiddelde aoi-lengte (189 karakters).

Hiermee is de leessnelheid van zesdeklassers duidelijk hoger dan die van tweedeklassers ($e^{(8.286 + .386)} = 5831$ ms), maar ze verschilt niet van de leessnelheid van vierdeklassers ($e^{(8.286 - .031)} = 3874$ ms). Ook vierdeklassers lezen gemiddeld beduidend sneller dan tweedeklassers. Daarmee kunnen we hypothese 1a ('oudere vwo-leerlingen lezen sneller dan jongere vwo-leerlingen') deels bevestigen.

Vanzelfsprekend is de leestijd afhankelijk van de lengte van een aoi. Per tien karakters meer neemt de leestijd in eerste instantie toe, maar naarmate aoi's langer worden, neemt de leestijd juist opvallend af. De relatie met zinslengte is $(.010 * N_{lengte\ aoi} - .003 * N_{lengte\ aoi}^2)$. Die afname in leestijd ontstond doordat bij langere aoi's vaker re-



Figuur 3.1 De first-pass dwell time van niet-kernzin-aoi's, kernzin-aoi's en aoi's van voorop geplaatste kernzinnen, afhankelijk van de aoi-lengte (aantal karakters), voor vwo 6, in seconden



Figuur 3.2 De totale kijktijd van kernzin-aoi's, niet-kernzin-aoi's en aoi's van voorop geplaatste kernzinnen, afhankelijk van de aoi-lengte (aantal karakters), voor vwo 6, in seconden

gressies optraden naar eerdere aoi's, waardoor de *first-pass*-waarde relatief laag is.

Wat kernzinnen betreft: aan betrekkelijk korte kernzinnen wordt meer leestijd besteed dan aan dito niet-kernzinnen, maar bij langere kernzinnen verdwijnt het onderscheid. Voor kernzinnen is de relatie $(.010 - .012) * N_{\text{lengte aoi}} - .003 * N_{\text{lengte aoi}}^2$. De kernzinnen die voorop in de alinea zijn geplaatst, worden duidelijk langer gelezen dan niet-kernzinnen.

Leestijden totale kijktijd. In Figuur 3.2 zijn de leestijden van niet-kernzinnen en beide typen kernzinnen weergegeven voor de totale kijktijd. Zesdeklassers gebruiken gemiddeld $e^{8.561}$ (= 5224) milliseconden voor het lezen van een niet-kernzin van gemiddelde lengte (189 karakters). Daarmee lezen vwo 6-leerlingen beduidend sneller dan vwo 2-leerlingen ($e^{(8.561 + .253)}$) = 6721 ms) maar niet sneller of langzamer dan vwo 4-leerlingen.

De leestijd is weer duidelijk afhankelijk van de aoi-lengte: naarmate een aoi langer wordt, stijgt de leestijd sterk. Deze toename vlakt echter af bij heel lange niet-kernzinnen. Zie Figuur 3.2. De relatie met zinslengte is $(.109 * N_{\text{lengte aoi}} - .003 * N_{\text{lengte aoi}}^2)$.

Een kernzin van gemiddelde lengte wordt over het algemeen nauwelijks langer ($e^{(8.561 + .035)} = 5410$ ms) bekeken dan niet-kernzinnen van gemiddelde lengte ($e^{8.561} = 5223$ ms), al is het onderscheid tussen beide wel significant. Naarmate een kernzin langer dan gemiddeld wordt, verdwijnt het onderscheid met een niet-kernzin. Kernzinnen die vooraan in de alinea zijn geplaatst, worden wel duidelijk consequent langer bekeken ($e^{(8.561 + .035 + .157)} = 6330$ ms, bij een gemiddelde lengte) dan niet-kernzinnen. Voor kernzinnen is de relatie $((.109 - 0.006) * N_{lengte\ aoi} - .003 * N_{lengte\ aoi}^2)$.

Regressies en varianties. In hoeverre keken leerlingen vaak terug naar eerdere zinnen of tekstdelen? Opvallend in Tabel 3.3 is dat de waarden in de totale kijktijd van de kortere aoi's relatief dicht bij de waarden in de *first-pass*-maat liggen. Wanneer voor alle jaarlagen de correlatie tussen beide leesmaten wordt bekeken, dan blijkt deze opvallend hoog ($r = .71$; $p < .001$): van niet-lineair leesgedrag tussen aoi's lijkt over het algemeen beperkt sprake. Beide leesmaten lijken erg sterk geassocieerd. Als deze correlatie wordt geattenuëerd voor onbetrouwbaarheid, dan wijken de twee leesmaten zelfs niet van elkaar af ($r = 1.0$). Als daarnaast het percentage regressietijd binnen de totale kijktijd per aoi wordt berekend – dat deel van de totale kijktijd van een aoi dat ontstond door terugkijken –, dan blijkt 9% van de totale kijktijd van een aoi te zijn besteed aan terugkijken: een laag percentage. Leerlingen bliken ogenschijnlijk dus over het algemeen weinig terug.

Als hierbij vervolgens de jaarlagen nader worden vergeleken, dan blijkt de correlatie tussen de *first-pass*-maat en de totale kijktijd voor vwo 2-leerlingen opvallend hoog ($r = .81$; $p < .001$). De correlatie voor vwo 4-leerlingen ligt in vergelijking daarmee lager ($r = .65$; $p < .001$), evenals die voor vwo 6-leerlingen ($r = .69$; $p < .001$). Het verschil van vwo 2 met vwo 6 en met vwo 4 is daarmee significant ($z \geq 5.09$; $p < .001$); dat tussen vwo 4 en 6 echter niet ($z = -1.04$; $p = .30$).

Het aandeel regressietijd binnen de totale kijktijd verschilt eveneens tussen jaarlagen ($F(2, 2515) = 4.96$; $p < .01$), waarbij enkel het onderscheid tussen vwo 2 en vwo 4 significant is ($p < .01$). Vwo 4-leerlingen kijken 11% van de totale leestijd terug naar aoi's, 4% meer dan vwo 2-leerlingen (die 7% van de totale kijktijd eraan besteden). Vwo 6-leerlingen besteden 9% van hun tijd hieraan. Hypothese 1c), 'oudere vwo-leerlingen lezen minder lineair dan jongere vwo-leerlingen', kunnen we dus bevestigen, al lijkt de omvang van dit niet-lineaire lezen gering.

Kijkt men dan ten slotte onder in Tabel 3.3 nog naar de varianties, dan blijkt vooral bij *first-pass dwell time* de residuele variantie zeer groot. Deze variantie, die bestaat uit de interactie tussen leerling en tekst, de interactie tussen aoi's en leerlingen en een gedeelte error, neemt beduidend af bij de totale kijktijd. Deze afname hangt hoogstwaarschijnlijk voor een belangrijk deel samen met het deels onsystematische karakter van oogbewegingen, dat zich vooral voordoet in de *first-pass*-maat: regelmatig keerden leerlingen na twee, drie fixaties in een nieuwe aoi weer terug naar de laatste woorden van de vorige aoi. De twee andere varianties gedragen zich bij beide leesmaten vergelijkbaar: het aandeel variantie binnen leerlingen is behoorlijk, terwijl het aandeel variantie dat toegeschreven kan worden aan de teksten, opvallend beperkt is.

3.3 Beantwoorden

Het antwoord. Als het gemiddelde totaal aantal goede antwoorden per jaarlaag wordt bekeken, dan blijken vwo 2-leerlingen gemiddeld 7 van de 18 vragen goed te hebben ($sd = 2.5$), vwo 4-leerlingen 9.6 vragen ($sd = 2.2$) en vwo 6-leerlingen 11.3 vragen ($sd = 2.2$). Er is daarmee een duidelijk onderscheid tussen jaarlagen ($F(2, 51) = 15.95; p < .001$), waarbij zowel het onderscheid tussen vwo 2 en vwo 4/6 significant is ($p < .001$) als dat tussen vwo 4 en 6 ($p = .05$). Daarmee kunnen we dus hypothese 2a ('oudere leerlingen hebben vaker een goed antwoord') bevestigen.

Tabel 3.4 Fit en passingsvergelijking van tien modellen voor de (logaritme van de) totale kijktijd (-2LL: -2log likelihood) bij het beantwoorden

Model	-2LL	Modelvergelijking			
		Modellen	$\Delta\chi^2$	Δdf	p
1 intercept	25563.41				
2 + lengte aoi	25527.79	2 vs 1	35.62	1	<.001
3 + aoi-type	22200.50	3 vs 2	3327.29	2	<.001
4 + aoi-type * lengte aoi	22187.60	4 vs 3	12.93	2	.001
5 + jaarlaag	22180.70	5 vs 4	6.84	2	.03
6 + jaarlaag * aoi-type	22141.34	6 vs 5	39.36	4	<.001
7 + jaarlaag * lengte aoi	22141.05	7 vs 6	.29	2	.87
8 + antwoord correct	22119.13	8 vs 7	21.92	1	<.001
9 + antwoord correct * aoi-type	22038.36	9 vs 8	80.77	2	<.001
10 + antwoord correct * jaarlaag	22037.23	10 vs 9	1.13	2	.57

Passingsvergelijking. Voor de beschrijving van het beantwoorden werd alleen de totale kijktijd in de analyse betrokken. Uit Tabel 3.4 blijkt dat het verdisconteren van de lineaire lengte van de antwoord-aoi's leidt tot passingsverbetering (Model 2 vs Model 1: $\Delta\chi^2(1) = 35.62; p < .001$). Als de drie verschillende aoi-typen (vraag-aoi, antwoord-relevante aoi en antwoord-irrelevante aoi) in leestijd verschillen, leidt dit tot verbetering (Model 3 vs Model 2: $\Delta\chi^2(2) = 3327.29; p < .001$), evenals wanneer de aoi-lengte van de drie verschillende aoi-typen wordt meegenomen (Model 4 vs Model 3: $\Delta\chi^2(2) = 12.93; p = .001$). Een model dat bekijkt in hoeverre leestijden verschillen tussen jaarlagen, heeft een betere passing (Model 5 vs Model 4: $\Delta\chi^2(2) = 6.84; p = .03$). Bij nadere bestudering van de schattingen bij dit model blijken vwo 6-leerlingen (net) niet significant sneller te antwoorden dan vwo 2-leerlingen ($\beta = -.147; p = .06$), noch sneller of langzamer dan vwo 4-leerlingen ($p = .44$), maar vwo 4-leerlingen gemiddeld sneller te antwoorden dan vwo 2-leerlingen ($\beta = -.207; p = .01$), waarmee we hypothese 2b ('oudere lezers antwoorden sneller') deels kunnen bevestigen. Als we de leestijden van jaarlagen vervolgens laten verschillen tussen de drie aoi-typen, bewerkstelligt dit evenwel een zeer duidelijke verbetering (Model 6 vs Model 5: $\Delta\chi^2(4) = 39.29; p < .001$). Een model dat de interactie tussen jaarlagen en aoi-lengte meeneemt, doet dit niet (Model 7 vs Model 6: $\Delta\chi^2(2) = .29; p = .87$). Bij de vervolgmogelijkheden, die de invloed van het geven van een goed of fout antwoord op

de leestijd onderzoeken, blijkt telkens een significante verbetering, behalve bij het interactie-effect met jaarlaag (Model 10 vs Model 9: $\Delta\chi^2(2) = 1.13; p = .57$).³

Leestijden aoi's antwoordproces. Tabel 3.5 geeft de parameterschattingen van de leestijden in totale kijktijd, gebaseerd op model 9, waarbij niet-significante parameters uit de analyse zijn gelaten (model 7: het interactie-effect tussen jaarlaag en aoi-lengte). Wanneer allereerst het beantwoorden in het algemeen wordt bekeken, dan blijken zesdeklassers van de drie aoi-typen relatief het langst naar de vraag te kijken (bij een gem. aoi-lengte: $e^{9.646} = 15460$ ms). Ze kijken vervolgens het langst naar een antwoord-relevante aoi (bij gem. lengte: $e^{8.320} = 4105$ ms). Naar een niet-relevante aoi kijken deze leerlingen het kortst (bij gem. lengte: $e^{7.356} = 1566$ ms). Vwo 6 kijkt dus duidelijk langer (bij gem. lengte: $e^{(8.320)} - e^{(7.356)} = 2540$ ms) naar een antwoord-relevante aoi dan naar een antwoord-irrelevante aoi.

Naarmate de vraagstelling in lengte toeneemt, kijken leerlingen duidelijk langer naar de vraag ($.057 * N_{\text{lengte aoi}}$). Naarmate antwoord-relevante aoi's in lengte toenemen, besteden leerlingen er iets minder totale kijktijd aan ($-.023 * N_{\text{lengte aoi}}$); hetzelfde geldt voor de antwoord-irrelevante aoi's ($-.020 * N_{\text{lengte aoi}}$). Het verschil in leestijd tussen vraag en antwoord-(ir)relevante aoi's wordt dus groter naarmate hun lengte toeneemt.

Tweedeklassers kijken opvallend langer naar de vraag (bij gem. lengte: $e^{(9.646 + .336)} = 21634$ ms) dan zesdeklassers. Ze kijken duidelijk korter naar een antwoord-relevante aoi (bij gem. lengte: $e^{(8.320 - .326)} = 2963$ ms) en ook korter naar een antwoord-irrelevante aoi (bij gem. lengte: $e^{(7.356 - .260)} = 1207$ ms) dan zesdeklassers. Vwo 2 besteedt dus meer aandacht aan de vraag en minder aan de tekst.

Vierdeklassers kijken niet significant korter of langer naar een vraag dan zesdeklassers (bij gem. lengte: $e^{(9.646 + .051)} = 16269$ ms). Ze wijden niet significant meer of minder kijktijd aan een antwoord-relevante aoi (bij gem. lengte: $e^{(8.320 + .041)} = 4277$ ms), maar kijken wel korter dan zesdeklassers naar een niet-relevante tekst-aoi (bij gem. lengte: $e^{(7.356 - .267)} = 1199$ ms).

Wat de relatie met het uiteindelijke antwoord betreft: leerlingen met een goed antwoord kijken over het algemeen duidelijk korter naar een vraag dan leerlingen met een onjuist antwoord (bij gem. lengte: voor vwo 6: $e^{9.646} - e^{(9.646 - .204)} = 2853$ ms korter). Leerlingen die een correct antwoord formuleren, bekijken duidelijk langer een antwoord-relevante aoi (voor vwo 6, bij gem. lengte: $e^{(8.320 + .478)} - e^{8.320} = 2516$ ms langer), en wat korter een antwoord-irrelevante aoi (voor vwo 6, bij gem. lengte: $e^{7.356} - e^{(7.356 - .162)} = 234$ ms korter).

3 Ook verdere interactie-effecten met jaarlaag leiden niet tot een significant betere passing. Nemen we bijvoorbeeld, in aansluiting op de modellen in Tabel 3.4, als model 11 het interactie-effect van het geven van een goed antwoord met jaarlaag en de drie antwoord-aoi-typen, voor de totale kijktijd, dan vinden we: Model 11 vs Model 10: $\Delta\chi^2(4) = 3.89; p = .42$. Met andere woorden: het geven van een goed dan wel fout antwoord staat ons hier niet verder toe verschillen aan te tonen tussen de jongere en oudere vwo-leerlingen.

Tabel 3.5 Parameterschattingen voor de (logaritme van de) totale kijktijd naar aoi's bij het beantwoorden, per jaarlaag

Parameter	coëfficiënt	(se)
vraag-aoi (vwo 6)	9.646	(.107)
antwoord-relevante aoi (vwo 6)	8.320	(.087)
antwoord-irrelevante aoi (vwo 6)	7.356	(.105)
- vraag-aoi * aoi-lengte ^a	.057	(.007)
- antwoord-relevante aoi * aoi-lengte ^a	-.023	(.008)
- antwoord-irrelevante aoi * aoi-lengte ^a	-.020	(.008)
Jaarlaag		
Δ vwo 2 * vraag-aoi	.336	(.108)
Δ vwo 2 * antwoord-relevante aoi	-.326	(.108)
Δ vwo 2 * antwoord-irrelevante aoi	-.260	(.099)
Δ vwo 4 * vraag-aoi	.051	(.051)
Δ vwo 4 * antwoord-relevante aoi	.041	(.108)
Δ vwo 4 * antwoord-irrelevante aoi	-.267	(.099)
Antwoord		
Δ antwoord correct * vraag-aoi	-.204	(.073)
Δ antwoord correct * antwoord-relevante aoi	.478	(.092)
Δ antwoord correct * antwoord-irrelevante aoi	-.162	(.083)
Varianties		
S ² tekst	.049	(.016)
S ² leerling	.040	(.009)
S ² residu	1.087	(.018)

^a lengte: gecentreerd rond gemiddelde aantal karakters (189 karakters) en gedeeld door 10

Proportie bekeken aoi's antwoordproces. In bovenstaande werd echter niet verdisconteerd dat leerlingen per vraag vaker naar een antwoord-irrelevante of antwoord-relevante aoi hebben kunnen kijken. Om te bekijken of de proportie bekeken aoi's varieert tussen jaarlagen, pasten we een multinomiaal model toe, waarin we de lengte van aoi's verdisconteerden en jaarlaag als hoofdeffect meenamen.

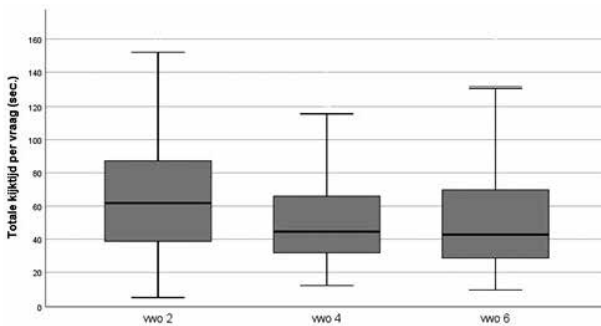
Dan blijkt dat er weliswaar over het algemeen tussen vwo 2 en vwo 6 een verschil aantoonbaar is ($F(2, 4962) = 3.65; p = .03$), maar niet tussen vwo 2 en vwo 4 ($F(2, 4962) = 1.78; p = .17$), en evenmin tussen vwo 4 en vwo 6 ($F(2, 4962) = 0.30; p = .74$). Als nader wordt ingezoomd op de afzonderlijke aoi-typen, dan zijn er voor de proportie antwoord-relevante aoi's geen significante verschillen tussen jaarlagen te vinden, maar wel voor de proportie antwoord-irrelevante aoi's. De kans dat een vwo 6'er er naar een voor de vraag irrelevante tekst-aoi kijkt, is heel wat kleiner ($\beta = -.341; p = .01$) dan de kans dat een vwo 2'er dat doet (bij een gem. aoi-lengte: $\beta = .399$). Een vierdeklasser heeft echter geen kleinere kans naar een irrelevante aoi te kijken dan een tweedeklasser ($\beta = -.259; p = .07$). Vwo 6 kijkt dus gemiddeld weliswaar langer naar een irrelevant tekstdeel dan vwo 2, ze kijkt wel duidelijk minder vaak naar deze delen. Uitgaande van een gemiddeld lange antwoord-irrelevante

aoi (195 karakters), dan bestaat 32% van de door een zesdeklasser bekeken aoi's uit voor beantwoording irrelevante tekstdelen, terwijl dat voor een tweedeklasser 61% is. Gevolg is dat de totale tijd die vwo 6 aan antwoord-irrelevante tekstdelen besteedt, aanzienlijk korter is, en dat vwo 6 daarmee het kortst naar de antwoord-irrelevante tekstdelen kijkt.

Nemen we in dit model ook mee of een leerling een goed dan wel fout antwoord had, dan blijkt dit duidelijk significant ($F(2, 4966) = 9.65; p < .001$). Als leerlingen een goed antwoord hebben, is de kans veel kleiner dat ze naar antwoord-irrelevante aoi's kijken ($\beta = -.444; p < .001$) dan wanneer ze een fout antwoord hebben. Door een vrij sterk interactie-effect met de aoi-lengte ($\beta = .137; p = .001$) hebben vooral de kortere antwoord-irrelevante aoi's aanmerkelijk minder kans bekeken te worden. Uitgaande van een gemiddeld lange antwoord-irrelevante aoi (195 karakters), bestaat 55% van de aoi's die een leerling met een goed antwoord bekijkt, uit voor beantwoording irrelevante tekstdelen terwijl dat voor een leerling met een fout antwoord 60% is.

Wanneer ten slotte de totaal bestede tijd aan een vraag wordt bekeken, dan blijken leerlingen ruim de tijd te nemen voor het lezen en beantwoorden van vragen. De gemiddeld bestede totale kijktijd voor het lezen en beantwoorden van een vraag was 60.5 seconden ($sd = 41.5$): ter vergelijking, de gemiddelde totale kijktijd voor het vooraf-lezen van een tekst bedroeg 86 seconden ($sd = 25.9$). In Figuur 3.3 is de frequentieverdeling van de totaal bestede tijd per vraagitem weergegeven per jaarlaag. Zoals al eerder vastgesteld, is enkel het onderscheid tussen vwo 2 en vwo 4 hier significant ($p = .01$); dat tussen vwo 2 en vwo 6 (nipt) niet ($p = .06$). In het leeuwendeel van de observaties blijken leerlingen aanzienlijk tijd nodig te hebben om te antwoorden, zeker als men zich realiseert dat leerlingen het antwoord niet op hoefden te schrijven en mondeling konden formuleren.

Varianties. Onderaan in Tabel 3.5 staan, ten slotte, de varianties. De residuele variantie, allereerst, is opmerkelijk hoog. Dat is niet bevreemdend. De hoge residuele variantie neemt het uiterst wisselende gedrag per vraag per persoon mee. De ene vraag is moeilijker dan de andere en veroorzaakt onder andere daardoor telkens



Figuur 3.3 Frequentieverdeling van de totale kijktijd per vraagitem, per jaarlaag, uitgedrukt in seconden

wisselend leesgedrag voor de vraag en voor de antwoord-relevante aoi's en antwoord-irrelevante aoi's. Daarmee vergeleken is de variantie die toegeschreven kan worden aan leerlingen gering, en dat geldt eveneens voor de variantie veroorzaakt door de vier sets van tekst en vragen.

Samenvattend: de verschillen tussen jaarlagen zijn hier voor een belangrijk deel conform verwachting. We zien allereerst hypothese 2c bevestigd: vwo 4 en vwo 6 kijken langer naar de antwoord-relevante aoi's dan vwo 2. Hypothese 2d zien we eveneens bevestigd, zij het via een omweg: zesdeklassers kijken weliswaar langer naar een niet-relevante aoi dan tweedeklassers, maar ze kijken er wel duidelijk minder vaak naar dan tweedeklassers, waardoor ze in totaal juist korter naar irrelevante tekstdelen kijken. Voorts: vwo 4 en vwo 6 kijken korter naar de vraag dan vwo 2. Wat vwo 6-leerlingen bij de leestijden van antwoord-aoi's onderscheidt van vwo 2-leerlingen, onderscheidt *grosso modo* ook de leerlingen met een goed antwoord van de leerlingen met een fout antwoord. En leerlingen besteden over het algemeen veel tijd aan een vraag.

3.4 Relatie tussen vooraf-lezen en beantwoorden

Om de relatie tussen het vooraf-lezen en het beantwoorden te verkennen, is een extra model opgesteld waarbij naast de gemiddelde varianties op leerlingniveau de covarianties tussen de vijf aoi-typen zijn geschat: aan de ene kant de kernzinnen en niet-kernzinnen bij het vooraf-lezen, en aan de andere kant de drie aoi-typen bij het beantwoorden. Daarbij hebben we wederom gecontroleerd voor aoi-lengte.

Leerlingen die bij het vooraf-lezen kernzinnen snel dan wel langzaam lezen, lezen ook de niet-kernzinnen snel dan wel langzaam ($r = 1.0$). Als leerlingen bij het beantwoorden een vraag snel/langzaam lezen, lezen ze bij het beantwoorden ook een antwoord-relevante of een antwoord-irrelevant tekstdeel snel/langzaam ($r \geq .66$). En wie een antwoord-relevant tekstdeel snel/langzaam leest, leest ook een antwoord-irrelevant tekstdeel snel/langzaam ($r = .80$). We vinden een minder hoog, matig verband tussen het lezen van de vraag en het vooraf-lezen van kernzinnen ($r = .65$) en niet-kernzinnen ($r = .51$): wie kernzinnen en niet-kernzinnen snel leest, leest ook de vraag snel. We vinden dus vanuit de leessnelheid van een leerling over het algemeen redelijk hoge correlaties tussen deze vijf factoren.

Is er ook een meer inhoudelijke samenhang aantoonbaar tussen vooraf-lezen en beantwoorden? Een leerling die bij het beantwoorden korter of langer naar een antwoord-irrelevante tekstdeel kijkt, heeft eerder ook enigszins korter of langer naar een kernzin gekeken ($r = .45$) én naar een niet-kernzin ($r = .37$) – een zwak verband. De relatie tussen het korter of langer kijken naar een antwoord-relevant tekstdeel bij het beantwoorden en het lezen van een kernzin en van een niet-kernzin bij het vooraf-lezen is echter zeer zwak, zelfs verwaarloosbaar (resp. $r = .17$ & $r = .02$). Door de relatief zwakke relaties blijken deze effecten niet te variëren tussen jaarlagen.

We kunnen hier dus niet of nauwelijks aantonen dat leerlingen die bij het vooraf-lezen langer kijken naar kernzinnen, ook langer naar de antwoord-relevante

tekstdelen kijken bij het antwoorden (hypothese 3a); wel kunnen we voorzichtig stellen dat leerlingen die bij het vooraf-lezen langer kijken naar kernzinnen, bij het beantwoorden wat korter naar antwoord-irrelevante tekstdelen kijken (hypothese 3b): al is het verband zwak en de hoeveelheid verklaarde variantie hierbij niet hoog (20%). Deze effecten blijken niet te variëren tussen jaarlagen (hypothese 3d).

3.5 Relatie tussen vooraf-lezen en het antwoord

Voor het verband tussen het vooraf-lezen en het antwoord (het aantal goede antwoorden) zijn weer beide leesmaten geanalyseerd: *first-pass dwell time* en totale kijktijd. Uit Tabel 3.6 blijkt dat de passing van een model waarin de proportie goede antwoorden wordt meegenomen, bij beide leesmaten niet verschilt van het basismodel (Model 2 vs Model 1: $\Delta\chi^2(1) \leq 1.77$; $p \geq .18$). Toevoeging van het interactie-effect tussen (niet-)kernzinnen en de proportie goede antwoorden levert evenmin verbetering op (Model 3 vs Model 2: $\Delta\chi^2(1) \leq 1.35$; $p \geq .25$). En een model dat het effect tussen (niet-)kernzinnen en de proportie correcte antwoorden laat variëren tussen jaarlagen, genereert evenmin een betere *fit* (Model 4 vs Model 3: $\Delta\chi^2(2) \leq 2.13$; $p \geq .35$).

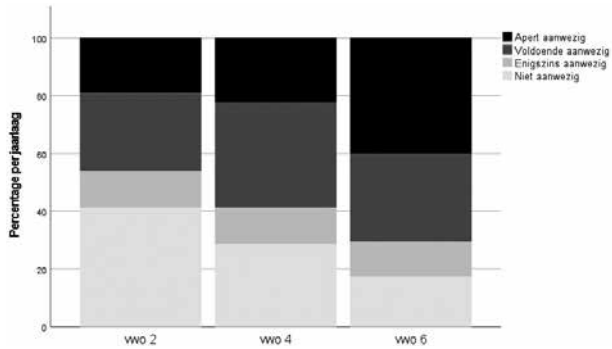
Hypothese 3c ('Leerlingen die langer bij het vooraf-lezen naar kernzinnen kijken, hebben bij de vragen vaker een goed antwoord') kunnen we dus vanuit beide leesmaten niet aantonen, en evenmin hypothese 3d ('Bij oudere leerlingen is dit duidelijker aantoonbaar').

Tabel 3.6 Fit en passingsvergelijking van drie modellen voor de (logaritme van de) *first-pass dwell time* en totale kijktijd (-2LL: -2log likelihood) voor de relatie tussen vooraf-lezen en het antwoord

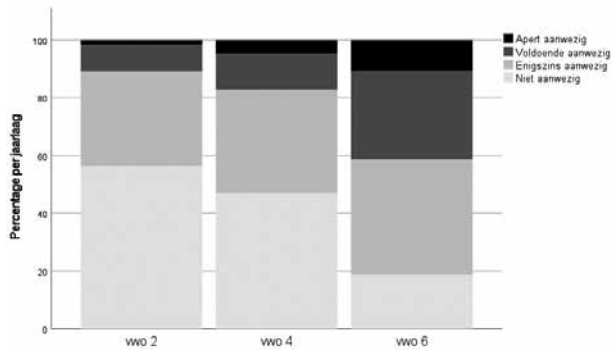
Model	-2LL	Modelvergelijking			
		Modellen	$\Delta\chi^2$	Δdf	p
<i>First-pass dwell time</i>					
1 basismodel	6356.54				
2 + proportie correct	6356.46	2 vs 1	.08	1	.78
3 + proportie correct * (niet-)kernzin	6355.11	3 vs 2	1.35	1	.25
4 + proportie correct * (niet-)kernzin * jaarlaag	6354.67	4 vs 3	.44	2	.80
<i>Totale kijktijd</i>					
1 basismodel	2804.73				
2 + proportie correct	2802.96	2 vs 1	1.77	1	.18
3 + proportie correct * (niet-)kernzin	2802.09	3 vs 2	.87	1	.35
4 + proportie correct * (niet-)kernzin * jaarlaag	2799.96	4 vs 3	2.13	2	.35

3.6 Retrospectief interview

In het interview zagen deelnemers hun oogbewegingsgedrag terug bij de beantwoording van de vier vraagitems bij de laatste tekst. We vroegen hun aan te geven in hoeverre ze meteen na het lezen van de vraag a) wisten in welke tekstpassages het antwoord te vinden was (tekstbasis), en b) een antwoord in gedachten hadden (situatiemodel). Bij vier leerlingen (twee tweedeklassers, een vierdeklasser, een



Figuur 3.4 Percentuele verdeling van de vier antwoordcategorieën in de tekstbasisscores, per jaarlaag



Figuur 3.5 Percentuele verdeling van de vier antwoordcategorieën in de situatiemodelscores, per jaarlaag

zesdeklasser) bleek bij één item de reflectie afwezig of om technische redenen onbruikbaar; bij één tweedeklasser gold dit voor drie items.

Over het algemeen was de audioreflectie, mede door de gelijktijdig afgespeelde scanpath, eenduidig en helder. Een beperkt aantal reflecties (twaalf gevallen, bij zes leerlingen) bleek onvoldoende eenduidig; deze gevallen werden voorgelegd aan twee leden van het screeningspanel. De drie codeurs stemden in tien gevallen overeen in hun codering. De twee overige gevallen werden buiten de analyse gelaten.

Worden dan allereerst de tekstbasis-scores bekeken, dan blijkt er een aantoonbaar verschil tussen de drie jaarlagen ($\chi^2(6) = 14.31; p = .03$). Het onderscheid tussen vwo 2 en vwo 6 is significant ($\chi^2(3) = 12.50; p \leq .01$) maar niet dat tussen vwo 2 en vwo 4, noch dat tussen vwo 4 en vwo 6 ($\chi^2(3) < 5.68; p \geq .07$).

In Figuur 3.4 is per jaarlaag de percentuele spreiding in tekstbasisscores weergegeven. Waar tweedeklassers in ruim 50% van de observaties aangeven niet/nauwelijks over een adequate tekstbasis bij beantwoording te beschikken, geven zesdeklas-

sers hiervan blijkt in een kleine 30% van de observaties. Tweedeklassers geven in 19% van de observaties aan over een apert aanwezige tekstbasis te beschikken, waar het bij zesdeklassers 40% betreft. De vwo 4-deelnemers positioneren zich tussen de scores van deze beide groepen in, al valt ook bij hen de geringere hoeveelheid 'apert aanwezig'-scores op in vergelijking met vwo 6. Vwo 6-leerlingen rapporteerden dus in het algemeen een aanzienlijk adequatere tekstbasis dan vwo 2-leerlingen.

Wanneer vervolgens de situatiemodel-scores per jaarlaag worden vergeleken, dan blijkt er wederom een apert verschil tussen jaarlagen ($\chi^2(6) = 29.55; p < .001$). Het verschil tussen vwo 2 en vwo 6 is significant, evenals dat tussen vwo 4 en vwo 6 ($\chi^2(3) \geq 15.50; p \leq .001$), maar niet dat tussen vwo 2 en 4 ($\chi^2(3) = 1.92; p = .30$).

In Figuur 3.5 is per jaarlaag de percentuele spreiding in situatiemodelscores weergegeven. Hier is te zien dat vwo 2 na vraaglezing in 56% van de observaties en vwo 4 in 47% van de observaties niet/nauwelijks van een adequaat situatiemodel blijkt gaf terwijl vwo 6 dit in 19% van de observaties aangaf. Er is ook een duidelijk verschil bij 'voldoende aanwezig': bij tweedeklassers betreft dit 9% van de scores en bij vierdeklassers 13%; bij zesdeklassers daarentegen betreft het 31%. Bij 'apert aanwezig' liggen de (lage) scores van de drie jaarlagen aanmerkelijk dichter bij elkaar. In veel opzichten geven vwo 6-leerlingen al met al blijkt van een aanmerkelijk sterker opgebouwd situatiemodel dan vwo 2- en vwo 4-leerlingen, maar ook zij hebben zelden meteen een scherp antwoord in hun hoofd.

4 Discussie en conclusie

4.1 Samenvatting

In deze studie onderzochten we verschillen in het vooraf-lezen van teksten en het beantwoorden van bijbehorende vragen tussen vwo 2-, vwo 4- en vwo 6-leerlingen. Daarbij generaliseren we over leerlingen én over teksten. We vatten allereerst de bevindingen samen voor het vooraf-lezen, het beantwoorden en voor de retrospectieve interviews.

We beginnen bij het vooraf-leesproces. Vwo 2-leerlingen lezen duidelijk langzamer dan vwo 6-leerlingen; vwo 4-leerlingen lezen sneller dan vwo 2-leerlingen, maar niet sneller of langzamer dan vwo 6-leerlingen (hypothese 1a). Deelnemers maken bij eerste lezing over het algemeen een zeker onderscheid tussen kernzinnen en niet-kernzinnen, maar bij het herlezen verdwijnt dit onderscheid vrijwel. We konden hierbij niet aantonen dat oudere meer aandacht hebben voor kernzinnen dan jongere leerlingen (hypothese 1b). Over het algemeen lazten deelnemers vooraf sterk lineair en ogenschijnlijk weinig strategisch. Tweedeklassers lezen een tekst meer lineair (minder strategisch) dan vierde- en zesdeklassers; vierdeklassers lezen echter niet meer of minder lineair (strategisch) dan zesdeklassers (hypothese 1c).

Betreffende het vragen beantwoorden: tweedeklassers hebben conform verwachting minder vragen goed dan de vierdeklassers, en de vierdeklassers weer minder vragen goed dan de zesdeklassers (hypothese 2a). Gemiddeld geldt: het

beantwoorden van een vraag neemt veel tijd in beslag, en slechts relatief zelden geeft men snel en adequaat antwoord op een vraag. Alleen voor vwo 4-leerlingen konden we aantonen dat ze over het algemeen sneller antwoord geven dan vwo 2-leerlingen (hypothese 2b). Leerlingen hebben bij het beantwoorden gemiddeld meer oog voor een antwoord-relevant tekstdeel dan voor een antwoord-irrelevant tekstdeel; de vraagstelling zelf krijgt duidelijk de meeste aandacht. Vierde- en zesdeklassers kijken langer naar een antwoord-relevant tekstdeel dan tweedeklassers. Maar zesdeklassers kijken niet langer naar een antwoord-relevant tekstdeel dan vierdeklassers; noch kijken zesdeklassers vaker ernaar dan tweede- en vierdeklassers (hypothese 2c). Wat de irrelevante tekstdelen betreft: hiernaar kijken zesdeklassers beduidend minder vaak dan tweedeklassers, maar ze kijken er niet minder vaak naar dan vierdeklassers; zesdeklassers kijken, onverwacht, ook juist langer naar een antwoord-irrelevant tekstdeel dan tweede- en vierdeklassers (hypothese 2d). Voor het overige: zesdeklassers kijken gemiddeld beduidend korter naar de vraag dan tweedeklassers.

Over het algemeen komt het onderscheid tussen oudere en jongere leerlingen overeen met het onderscheid tussen leerlingen met een goed of fout antwoord: zowel zesdeklassers als leerlingen met een goed antwoord kijken gemiddeld korter naar de vraag en gebruiken de tekst selectiever.

Wat de relatie tussen het lees- en het antwoordproces betreft: leerlingen die bij het vooraf-lezen langer kijken naar kernzinnen, kijken bij het antwoorden niet of nauwelijks langer naar de antwoord-relevante tekstdelen (hypothese 3a). Wel kunnen we voorzichtig stellen dat deze leerlingen bij het beantwoorden wat korter naar antwoord-irrelevante tekstdelen kijken (hypothese 3b), al is het aangetoonde verband zwak. We konden tussen jaarlagen bij het voorgaande geen onderscheid aantonen (hypothese 3d). Voorts: leerlingen die langer naar kernzinnen kijken, hebben bij het beantwoorden, onverwacht, niet aantoonbaar vaker een goed antwoord (hypothese 3c), waarbij jaarlaag – wederom – niet aantoonbaar een rol speelt (hypothese 3d). We kunnen vanuit de eye-track-data daarmee opvallend weinig relaties aantonen tussen vooraf-lezen en vragen beantwoorden (basishypothese A).

Dan het retrospectieve interview: leerlingen gaven daarin aan direct na vraagstelling over het algemeen over een redelijke tekstbasis te beschikken, maar in het grote merendeel van de gevallen niet over een redelijk tot goed situatiemodel (basishypothese A). De zesdeklassers beschikten daarbij over een beduidend betere tekstbasis dan de tweedeklassers; een onderscheid tussen vierdeklassers en tweed- dan wel zesdeklassers konden we hier niet aantonen. Vierdeklassers en zesdeklassers gaven vaker blijk van een adequaat situatiemodel dan de tweedeklassers, en de zesdeklassers weer vaker dan de vierdeklassers (basishypothese B).

Ten slotte: bij het vooraf-lezen valt het op dat de variantie tussen leerlingen aanmerkelijk hoger ligt dan de variantie tussen teksten: vooral de individuele leesaanpak van een leerling verklaart de variatie in kijkgedrag, de teksten zelf aanmerkelijk minder. Daarnaast: de residuele variantie is vooral bij het beantwoorden uiterst omvangrijk. Elke vraag lijkt vooral telkens wisselend, van persoon tot persoon sterk afwijkend antwoordgedrag te genereren.

4.2 *Discussie*

Het doel van deze studie is verschillen aan te tonen tussen jongere en oudere vwo-leerlingen bij de tekst met vragen. In zeer algemene lijnen tekenen in de resultaten de verschillen zich dan af zoals verwacht: zesdeklassers onderscheiden zich van tweedeklassers door een hogere leessnelheid, meer terugkijken (meer strategisch lezen) bij het doornemen van de teksten, en door een effectiever antwoordproces en adequater tekstbegrip bij het beantwoorden van de vragen. Tegelijkertijd treffen we ook onverwachte resultaten aan. De verschillen tussen jaarlagen bij vooraf-lezen én beantwoorden zijn vaak minder markant dan verwacht. Vierdeklassers gedragen zich vooral in de eye-track-resultaten vaak als zesdeklassers, met name bij het vooraf-lezen. Bovendien blijkt tussen het vooraf-lezen en het vragen beantwoorden een aanzienlijk minder sterke relatie dan verwacht. We lopen de opvallende bevindingen langs.

Kijkt men allereerst naar het vooraf-lezen, dan valt op dat de deelnemers, onafhankelijk van jaarlaag, tijdens de eerste doorlezing aan kernzinnen wat meer aandacht schenken. Dit bevestigt voorzichtig de aanname van de Expertgroep Doorlopende Leerlijnen Taal (2008, 2009) dat lezers vanaf niveau 2F hoofd- en bijzaak scheiden, én de belangrijke aanname in de Constructie-Integratie-theorie (Kintsch, 1998) dat geoefende lezers in eerste instantie onderscheid maken tussen hoofd- en bijzaken. Bij herlezing van zinnen verdwijnt echter onverwacht het onderscheid tussen kernzinnen en niet-kernzinnen. Vanuit de oplopende moeilijkheidsgraad van de stimuli verwachtten we bovendien dat de oudere leerlingen kernzinnen gemiddeld langer bekijken dan de jongere leerlingen, maar ook dit konden we niet aantonen. Is kernzindetectie bij jongere expert-lezers misschien niet afhankelijk van tekstcomplexiteit? Een andere, verleidelijke verklaring volgt uit de studie van Yeari et al. (2015). Zij vonden dat leerlingen zich bij een leestaak met open vragen vooral bij de herlezing op de hele tekst richten, niet op de hoofdzaken: ze kunnen immers over alle tekstelementen vragen krijgen. Speciaal aandacht schenken aan kernzinnen wordt niet door de taak beloond.

Daarnaast valt het op dat deelnemers ondanks de deels pittige teksten over het algemeen sterk lineair lezen, ook al vinden we wel voorzichtig de verwachte verschillen tussen jaarlagen. Frequente inzet van leesstrategieën lijkt over het algemeen onaannemelijk. De Expertgroep Doorlopende Leerlijnen Taal (2008, 2009) veronderstelt in de beschrijving van het domein zakelijk lezen 3F en 4F de adequate inzet van enkele leesstrategieën, evenals de Nederlandse eindtermen voor het vwo. Dit strategische leesgedrag hebben wij bij het vooraf-lezen nauwelijks kunnen aantonen. Daarmee is natuurlijk niet gezegd dat vwo-leerlingen deze strategieën niet kennen en kunnen toepassen. Eerder lijken ze intensieve inzet ervan bij dit experiment niet noodzakelijk te hebben geacht. Uit enkele recente verkennende studies naar het leesgedrag van vwo'ers bleek tamelijk oppervlakkig, sterk pragmatisch leesgedrag (zie hoofdstuk 2; Breukink, In voorbereiding; Stevens, 2018). In deze studie lijkt zich een soortgelijk oppervlakkig leesproces af te tekenen.

Ten derde valt op dat de zesdeklassers zich in de eye-track-resultaten vaak nauwelijks onderscheiden van de vierdeklassers. Dit kan meerdere verklaringen

hebben, bijvoorbeeld de samenstelling van de steekproef (veel goede lezers onder de vwo 4-leerlingen?). Eén verklaring ligt voor de hand: een zesdeklasser weet tijdens het lezen beter te infereren en expliciteren, zonder dat hij daarbij duidelijk anders op zinnen en woorden fixeert. Daarmee wordt op een bekende begrenzing van eye-tracking gestuit (Holmqvist & Andersson, 2017). De betere situatiemodelscores van de zesdeklassers bij het retrospectieve interview kunnen hiervoor als bewijs dienen.

Wat het vragen beantwoorden betreft, valt op dat zesdeklassers vaak adequater antwoordgedrag vertonen dan tweedeklassers en, in mindere mate, vierdeklassers. Zesdeklassers lijken het meest verfijnde antwoordgedrag te vertonen: ze doen niet korter over de vraagbeantwoording dan tweedeklassers, maar lezen de vraag veel korter en bekijken een antwoord-irrelevant tekstdeel minder vaak maar wel langer dan tweedeklassers. Daarmee hebben ze ook vaker vragen goed. Deze resultaten zijn logisch te verklaren: wie vooraf een redelijk beeld van de tekst heeft gekregen, begrijpt de vraagstelling sneller en weet beter waar het antwoord zich bevindt, waardoor hij zijn aandacht gericht en relatief ruimer kan wijden aan de tekstbestudering. En andersom: wie de vraag minder makkelijk of niet begrijpt, zal deze vaker moeten herlezen en richt daarmee minder de aandacht op de tekst, waardoor zij of hij al met al gemakkelijker een fout antwoord geeft. Ook Brunfaut en McCray (2015) vonden in hun eye-trackdata dat leerlingen met een goed antwoord sneller antwoordgedrag vertoonden, met minder aandacht voor de vraag, evenals Bax (2013) dat vond bij onderzochte testitems. Betere lezers antwoorden gericht en effectiever.

Wat de relatie tussen vooraf-lezen en beantwoorden betreft: opvallend genoeg konden we onze twee basishypothesen maar zeer ten dele bevestigen. We konden een redelijk sterk verband tussen het vooraf-lezen en het vragen beantwoorden aantonen op tekstbasisoniveau, maar niet op situatiemodelniveau. Desalniettemin rijst er een tamelijk eenduidig beeld op uit de resultaten: leerlingen bouwen tijdens het vooraf-lezen een nogal globaal, weinig specifiek situatiemodel op, dat wel wat in scherpte toeneemt naarmate de jaarlaag hoger is. Ondanks onze zeer nadrukkelijke instructie vooraf om de teksten intensief te bestuderen en de relatief hoge concentratie- en motivatiecijfers, volstonden leerlingen tijdens vooraf-lezen, lijkt het, met *shallow processing* (Sanford & Sturt, 2002). Vooral bij het beantwoorden van de vragen bouwden ze een accuraat en adequaat situatiemodel op. Dit correspondeert met eerdere bevindingen over het effect van het stellen van open vragen bij een tekst (bijv. Cerdán et al., 2009; Rouet et al., 2001), waarin de participanten door het beantwoorden van open vragen een diepgaand situatiemodel bij de tekst construeerden. Pas toen leerlingen door de vragen te lezen in detail duidelijk werd wat ze met de gelezen tekst moesten doen, zetten ze zich tot diepgaande bestudering. Daarmee lijkt vooral de beantwoordingsfase te bepalen of een leerling een goed dan wel fout antwoord geeft (vgl. hierover: Embretson & Wetzels, 1987; Rupp et al., 2006).

Meer dan vooraf verwacht, wijzen de aangetoonde verschillen tussen jaarlagen vaak op sterk taakafhankelijk lees- en antwoordgedrag bij de tekst met vragen. De vraag is in hoeverre dit sterk taakafhankelijke gedrag onverwacht is. Leesgedrag is

immers altijd sterk doelafhankelijk (Graesser et al., 1994), en het wordt vanuit onderzoek steeds duidelijker dat lezers vaak niet met volle, rijke aandacht lezen, maar hun leesgedrag zeer pragmatisch aanpassen aan de vereisten voor een taak (vgl. de *minimal task-elaboration hypothesis*: Britt et al., 2018). Bij de tekst met vragen wordt het doel voor leerlingen pas bij het lezen van de vragen duidelijk en dan is het niet zinvol tijdens het vooraf-lezen een accuraat, rijk gespecificeerd situatiemodel op te bouwen. Globaal de tekst lezen is echter voor een lezer wel zinnig: hij krijgt zo een beeld van de hele tekst en voorkomt daarmee de valkuil dat hij op basis van louter lokale informatie een vraag onjuist interpreteert of beantwoordt (vgl. Cerdán et al., 2009; Ozuru et al., 2007). De soms weinig markante verschillen tussen jaarlagen bij het vooraf-lezen in deze studie kunnen hiermee worden verklaard.

Hoezeer taakverwachting en -vertrouwdheid van leerlingen bij de resultaten in deze studie inderdaad een rol spelen, kan blijken uit een tot nog toe onbenoemd detail. Acht van de 52 participanten gaven bij het retrospectief interview ongevraagd (!) aan dat ze zich minstens eenmaal in hun antwoordgedrag volledig lieten leiden door conventies rondom de taak tekst met vragen, niet door enig tekstbegrip: “Ik zocht het antwoord in alinea 1, omdat dit de eerste vraag bij de tekst was.”

Dit pragmatische leesgedrag wordt waarschijnlijk nog in de hand gespeeld door de huidige inrichting van de tekst met vragen in het vak Nederlands. Men stelt meestal geen duidelijk leesdoel vooraf en een leerling kan sterk in aard variërende vragen verwachten (vragen op lokaal of globaal niveau; analytische, stilistische, interpreterende vragen), waarbij hogere-orde-denkvaardigheden als reflecteren en evalueren enkel beperkt aan bod komen. Vaak hebben de vragen ook betrekking op één tekst, en de taak dient doorgaans individueel uitgevoerd. Met deze conventies rond de tekst met vragen is een leerling al vertrouwd vanaf de lagere school en in bovenbouwmethodes Nederlands als *Nieuw Nederlands* en *Op niveau* domineert ze de leesvaardigheidsparagrafen (vgl. Van der Leeuw et al., 2017).

4.3 Kanttekeningen

Enkele algemene voorbehouden bij deze studie zijn op hun plaats. Het betreft primair een eye-trackstudie. De basisaanname achter deze methode, namelijk dat kijken naar een aoi een directe weerspiegeling is van de aandacht die een participant eraan schenkt, staat nog altijd ter discussie, evenals de praktische implicaties van deze aanname (vgl. Holmqvist & Andersson, 2017). Eye-tracking biedt daarnaast alleen zicht op het hoe, niet op het waarom achter lezen. En weliswaar vulden we deze lacune deels door een retrospectief interview in te zetten, hierin werden alleen de laatste vier vragen bij één tekst besproken. Daarnaast betrof het een cross-sectioneel, dus niet longitudinaal onderzoek, afgenomen op één Nederlandse school, bij slechts vier relatief korte teksten en een beperkt aantal vragen. Ten slotte: de verschillen tussen de drie jaarlagen onderzochten we door telkens groepen van minder dan twintig leerlingen met elkaar te vergelijken. De steekproeven bij vergelijkbare *mixed method*-eye-trackstudies zijn overigens vergelijkbaar of zelfs aanmerkelijk kleiner (vgl. Bax, 2013; Brunfaut & McCray, 2015).

Tegelijk wijzen we graag op het methodologisch vernieuwende karakter van deze studie. Hoe de mentale ‘complexen’ bij een tekst met vragen zich tot elkaar verhouden – initieel leesproces, antwoordproces en het uiteindelijke antwoord –, is methodologisch ingewikkeld te ontrafelen. Gebruik van hardop-denkprotocollen werd wel ingezet, maar grijpt sterk in in het natuurlijke lezen (bijv. Norris, 1990; vgl. Pressley & Afflerbach, 1995). Om de relatie tussen vooraf-lezen en vragen beantwoorden te onderzoeken, gebruikten Cerdán et al. (2009) daarom sterk op eye-tracking gelijkende software om het vooraf-lezen te beschrijven, maar zij stelden in hun studie tekstbegrip gelijk aan de score op de vragen. Enkele redelijk recente Britse studies combineerden eye-tracking met hardop-denken via retrospectieve interviews (Bax, 2013; Brunfaut & McCray, 2015), maar lieten de initiële tekstlezing buiten beeld. In onze studie wordt deze *mixed-method* benadering overgenomen om initieel leesgedrag en antwoordgedrag, afzonderlijk en in samenhang, te kunnen beschrijven. Door te generaliseren over leerlingen en teksten, konden we zo ‘holistisch’ verschillen in tekstbegrip tussen jongere en oudere vwo-leerlingen aantonen bij de taak tekst met vragen.

4.4 *Implicaties voor het onderwijs*

Al met al lijkt deze studie daarmee de vaak geuite kritiek bij de tekst met vragen te ondersteunen dat leerlingen met de tekst met vragen vooral vragen leren beantwoorden (bijv. Hoogeveen & Bonset, 1998). Eerdere Nederlandse studies vonden dat taken zoals de sorteertaak en de mental model-taak beter op het niveau van het situatiemodel tekstbegrip meten dan de tekst met vragen (Kamalski, 2007; Land, 2009). Vanuit deze studie is de vraag in hoeverre de tekst met vragen in zijn huidige inrichting leerlingen leert om zelfstandig, met dieper tekstbegrip teksten te lezen. Het is bepaald niet onwaarschijnlijk dat met de veelvuldige inzet van deze taak eerder taakvaardigheid wordt bevorderd dan leesvaardigheid.

De sterke dominantie van de tekst met vragen als individuele taak in de lessen Nederlands in zijn huidige vorm lijkt zo ongewenst. Recent klonken er in Nederland geluiden dat de motivatie van leerlingen om ‘diep te lezen’ afneemt (Onderwijsraad & Raad voor Cultuur, 2019) en er zijn tekenen dat het tekstbegrip in Nederland en Vlaanderen daalt (Van den Branden et al., 2019; Gubbels et al., 2019). De vraag is in hoeverre de huidige didactische dominantie van de tekst met vragen daaraan bijdraagt, een samenhang die in kranten nogal eens wordt verondersteld (bijv. Welgraven & De Vos, 2019). De kritiek op het Nederlandse lesonderwijs luidt van oudsher dat het sterk toetsgericht is (bijv. Hoogeveen & Bonset, 1998): leraren trainen hun leerlingen vooral die kennis en vaardigheden die nodig zijn om ‘teksten met vragen’ goed te kunnen maken. Daarmee ontstaat al snel het gevaar dat andere cognitieve vaardigheden, betrokken bij ‘diep lezen’, het onderspit delven: er ontstaat “survival of content that fittest best the high-stakes test” (Popham, 2001).

Het lijkt in die zin raadzaam om de dominantie van de ‘reguliere’ tekst met vragen te doorbreken. Dat kan door een veelheid aan andere leestaken aan te bieden: men

denke aan samenvattingstaken of synthesetaken (Van Ockenburg, 2022), argumentatietaken (Anmarkrud et al., 2014) of sorteertaken (Land, 2009). Ook een andere inrichting van de tekst met vragen is zeer te overwegen: tekstafhankelijke vragen kunnen steeds in groeps- of klassikaal verband worden behandeld (vgl. Fisher & Frey, 2015); hogere-orde-denkvaardigheden zoals evalueren en reflecteren kunnen meer een plaats vinden, en de taak kan voorzien worden van een duidelijk leesdoel, met aan dat leesdoel gekoppelde vragen (Nederlands nu! & Sectiebestuur Nederlands van Levende Talen, 2018; Schuurs, 2019). Bovenal lijkt het zinnig het proces achter het 'diep lezen' zelf centraal te blijven stellen, bijvoorbeeld door modellen (Fisher et al., 2008) of de inzet van observerend leren (Couzijn, 1995; Keehnen et al., 2015).

HOOFDSTUK 4

De relatie tussen het vooraf-lezen van teksten en het beantwoorden van begripsvragen

*Twee experimenten onder vwo 5-leerlingen**

* Dit hoofdstuk is in een eerdere vorm verschenen als: Rooijackers, P., Silfhout, G. van, Schuurs, U., & Bergh, H. van den. (2021). De relatie tussen het vooraf lezen van teksten en het maken van begripsvragen: twee experimenten onder 5-vwo-leerlingen. *Pedagogische Studiën*, 98(5), 320-349.

Deze studie werd door de Vereniging voor Onderwijsresearch (VOR), het Vlaams Forum voor Onderwijsresearch (VFO) en de tijdschriftredactie bekroond tot het beste artikel dat in 2021 in *Pedagogische Studiën* is verschenen.

Samenvatting

In Nederland traint men vaak tekstbegrip met gebruikmaking van teksten met vragen: leerlingen bestuderen vooraf een tekst en beantwoorden vervolgens tekstafhankelijke vragen, waarbij de tekst raadpleegbaar blijft. Daarbij wordt doorgaans verondersteld dat wie een tekst vooraf intensief bestudeert, de vragen vaker correct beantwoordt.

In deze studie is de relatie tussen initiële tekstbestudering en antwoordscore nader onderzocht, onder vwo 5-leerlingen, bij twee zakelijke teksten, met begripsvragen, via twee experimenten. Experiment I onderzocht de mate waarin de *beschikbaarheid* van de tekst tijdens vraagbeantwoording de antwoordscore beïnvloedt, nadat deelnemers ($N = 174$) vooraf teksten intensief hadden bestudeerd. Experiment II onderzocht in hoeverre de *intensiteit* van vooraf-tekstbestudering de antwoordscore beïnvloedt, als deelnemers ($N = 398$) bij de vragen de tekst kunnen raadplegen.

Resultaten: in experiment I beantwoorden leerlingen die bij de vragen de tekst wél kunnen raadplegen, de vragen beduidend vaker correct dan leerlingen die deze niet kunnen raadplegen. In experiment II kan niet worden aangetoond dat leerlingen die de tekst vooraf intensief bestuderen, hogere scores dan leerlingen die de tekst vooraf minder of niet-intensief bestudeerden. Het antwoordproces wordt kennelijk niet zozeer beïnvloed door intensiteit van vooraf-lezen, als wel door herlezen tijdens het antwoordproces. De vraag is daarmee in hoeverre deze taak leerlingen ertoe traint zelfstandig een tekst te bestuderen.

1 Inleiding

1.1 Aanleiding

In het schoolvak Nederlands wordt in de leesles doorgaans gewerkt met teksten met vragen: leerlingen bestuderen een tekst en beantwoorden vervolgens begripsvragen, waarbij de tekst raadpleegbaar blijft (Linthorst & De Glopper, 2015; SLO, 2016a). Op de nogal dominante positie van deze taak in de leesles Nederlands is de afgelopen decennia vaak door neerlandici en onderwijskundigen in Nederland en Vlaanderen kritiek geuit (bijv. Van den Branden, 2003; Hoogeveen, 2018; Hoogeveen & Bonset, 1998; Van de Wetering & Groenendijk, 2015). Het markantst uitte zich in Nederland deze kritiek rondom het Centraal Examen Nederlands havo/vwo, dat vooral gebruikmaakt van vragen bij teksten en dat volgens critici niet zozeer de leesvaardigheid van leerlingen zou toetsen als wel ‘vraagvaardigheid’, de mate waarin ‘je goed bent in examen doen’ (Grezel, 2014).

Een belangrijke veronderstelling achter de didactiek bij deze taak is doorgaans dat wie de tekst vooraf intensief bestudeert, de vragen vervolgens vaker correct beantwoordt – een veronderstelling die wortelt in opvattingen over tekstbegrip in de neerlandistiek uit de jaren tachtig van de vorige eeuw (vgl. Drop, 1979; Wesdorp, 1981; zie Rooijackers et al., 2021a). Deze ‘simpele’ relatie tussen de initiële tekstbestudering en vraagbeantwoording vinden we impliciet terug in oefenboeken voor het Centraal Schriftelijk Examen Nederlands (bijv. ExamenOverzicht, 2020) en in een aantal Nederlandse leergangen voor het schoolvak Nederlands. Zo vraagt de bovenbouweditie vwo van de in Nederland veelgebruikte leergang *Nieuw Nederlands* (6^e editie) van leerlingen vaak dat ze eerst een tekst via een uitgebreid “stappenplan tekstanalyse” doorwerken, voordat ze de vragen over de tekst gaan beantwoorden (Frank et al., 2019).

De empirische onderbouwing bij deze veronderstelling is echter schaars en onvolledig. Wanneer een onderscheid gemaakt wordt tussen wel en niet vooraf-lezen van de tekst, dan lijkt wél vooraf-lezen bij deze taak inderdaad te leiden tot hogere resultaten dan direct starten met de vragen (Cerdán et al., 2009; De Milliano & Van Gelderen, 2020; Salmerón et al., 2015). Maar daarmee is de vraag nog niet beantwoord of de *intensiteit* van de vooraf-bestudering ook van invloed is op de woordscore, zoals vaak verondersteld. Studies hiernaar ontbreken naar ons weten.

Deze veronderstelling spreekt in elk geval niet vanzelf. Vooral tijdens het antwoordproces blijken leerlingen een mentale representatie van (delen van) de tekst op te bouwen om antwoord te kunnen geven, en niet zozeer tijdens het vooraf-lezen (zie hoofdstuk 2 en 3; Schroeder, 2011): doordat leerlingen tijdens de vraagbeantwoording de tekst nog kunnen raadplegen, kunnen ze ontbrekend of inadequaat tekstbegrip repareren en hun antwoord controleren. Een intensieve vooraf-bestudering heeft voor leerlingen dan op het oog weinig meerwaarde. Die situatie verandert als leerlingen bij het beantwoorden van de vragen *niet* meer over de tekst kunnen beschikken: in zo'n taaksetting blijken hogere-ordedenkprocessen tijdens het vooraf-lezen, zoals conceptuele integratie en coherente representatie van de tekstinhoud, grotendeels verschillen in antwoordscore van leerlingen te kunnen verklaren (Schroeder, 2011; vgl. Ferrer et al., 2017; Ozuru et al., 2007). Doordat leerlingen niet langer ontbrekend of onvoldoende tekstbegrip tijdens de antwoordfase kunnen repareren, zijn ze gedwongen hun antwoord grotendeels te baseren op hun vooraf-opgebouwde mentale representatie van de tekst, en daarmee krijgt een intensieve vooraf-tekstbestudering in deze taaksetting een duidelijk meerwaarde. Deze onderzoeksresultaten houden echter niet in dat een intensieve vooraf-tekstbestudering tot meer correcte antwoorden leidt als leerlingen bij het beantwoorden van de vragen nog wél over de tekst kunnen beschikken.

In deze studie willen we daarom nagaan in hoeverre een intensieve vooraf-tekstbestudering tot een hogere tekstbegripsscore op de vragen leidt, zoals vaak bij het schoolvak Nederlands verondersteld. Deze veronderstelling vormt immers een belangrijk argument om de 'tekst met vragen' veelvuldig als oefenmateriaal in de les in te zetten. Om deze vraag te kunnen beantwoorden, hebben we in deze studie getracht de relatie tussen initieel leesproces en antwoordscore nader bloot te leggen. Dat doen we door twee factoren te manipuleren: a) de beschikbaarheid van de tekst bij het beantwoorden van de tekstafhankelijke vragen, en b) de intensiteit waarmee lezers vooraf een tekst bestuderen.

1.2 *Tekstbegrip en intentionaliteit*

Allereerst: wat is tekstbegrip? Een aantal breed geaccepteerde theoretische modellen van vóór de millenniumwisseling voorspellen en verklaren inmiddels de cognitieve processen tijdens tekstbegrip, zoals de Constructie-Integratie-Theorie (CI-Theorie: Kintsch, 1998; Kintsch & Van Dijk, 1978) en het Landschapsmodel (Van den Broek et al., 1999). Een belangrijk kernconcept hierin is het onderscheid tussen tekstbasis en situatiemodel. Een tekst wordt in deze theorieën beschouwd als een netwerk van hiërarchisch samenhangende, betekenisvolle beweringen. Hierin zijn lezers steeds op zoek naar samenhang, in eerste instantie tussen woorden en zinnen (microstructuur) en in tweede instanties tussen alinea's (macrostructuur). Al lezend bouwen ze zo gaandeweg een 'tekstbasis' op, een vrij letterlijke mentale representatie van de tekst, en een 'situatiemodel', een mentaal model waarin ze de tekstuele wereld integreren in hun eigen voorkennis, opvattingen en ideeën. Pas op dit laatste niveau ontstaat diepere betekenis: de lezer heeft zich de tekstinhoud eigengemaakt.

In deze twintigste-eeuwse theoretische modellen bleef nog goeddeels onuitgevoerd hoe specifieke leesdoelen in studie- en arbeidscontexten leesgedrag kunnen beïnvloeden, zoals in onze studie bij het beantwoorden van begripsvragen over een tekst. Sinds de millenniumwisseling is in internationaal leesonderzoek steeds meer deze intentionaliteit achter leesgedrag in kaart gebracht, de situationele en functionele inbedding van tekstbegrip. Uiteenlopende leesdoelen en leessituaties blijken bij lezers tot deels afwijkende begripsprocessen en een afwijkende inzet van vaardigheden en strategieën te leiden (bijv. Van den Broek & Helder, 2017; McCruden et al., 2011; Yeari et al., 2015). Al naargelang het doel dat lezers zichzelf stellen, bouwen ze een tekstuele representatie op die kwalitatief en kwantitatief afdoende lijkt om aan het doel van de taak te voldoen (bijv. Van den Broek et al., 1995). Geruggensteund door deze studies naar de intentionaliteit achter leesgedrag wil het recente RESOLV-model (Britt et al., 2018; zie ook Rouet et al., 2017) ook het antwoordgedrag bij leestaken verklaren en voorspellen, een dimensie van leesgedrag die in de modellen van vóór 2000 nog grotendeels onbesproken bleef.

Het RESOLV-model beschouwt lezen voor een taak (*REading*) als het oplossen van een probleem (*SOLVing*). Bij een taak over een of meerdere teksten vertrekken lezers daarom vanuit een statisch *contextmodel*, dat zaken verdisconteert als de taakinterpretatie, de status van de 'opdrachtgever', de hulpmiddelen en belemmeringen bij de taak, evenals een zelfbeoordeling bij deze taak door de lezers. Vanuit dit contextmodel stellen lezers vervolgens een dynamisch *taakmodel* op: taakspecifieke doelen en scenario's, waarin ze interpreteren wat ze moeten doen om de doelen te bereiken, en welke procedures ze daarbij moeten hanteren. In tegenstelling tot het contextmodel, dat in principe na het starten met de taak vastligt, is dit taakmodel dynamisch: lezers kunnen het gaandeweg de taak aanpassen en bijstellen. Afhankelijk van dit taakmodel bouwen ze van de tekst(en) een passende mentale representatie op en zetten ze zich tot een eventuele vervolgtask.

Het RESOLV-model staat ons in deze studie toe verwachtingen te formuleren over lees- en antwoordgedrag van lezers bij een tekst met vragen. Twee van de vijf basishypothesen bij dit model verdienen vooraf nog nadere uitwerking. Volgens de '*minimal task-elaboration hypothesis*' plannen lezers vooraf niet uitvoerig hoe ze een taak moeten aanpakken. Ze vertrekken weliswaar in hun taakmodel vanuit een hogere-orde-doel, maar ze werken vooraf enkel in zeer algemene lijnen uit hoe ze een taak willen aanpakken; al naargelang de uitvoeringsproblemen die ze ervaren, passen ze min of meer noodgedwongen gaandeweg hun werkwijze aan. Daarnaast: volgens de '*parsing the context hypothesis*' beslissen lezers vanuit hun contextmodel vooraf niet alleen óf ze de taak zullen uitvoeren, maar ook de mate waarin ze dit doe. Zo zullen lezers zich bij een zwaarwegende leestoets tot een diepere, nauwkeurigere en meer gelaagde werkwijze zetten dan bij een 'alledaagse' leestaak van gelijke strekking. In de volgende paragraaf werken we deze twee hypothesen nader uit door ze toe te passen op de taak 'tekst met vragen'.

1.3 Taakmodel bij een tekst met vragen

Vragen stellen bij een tekst is waarschijnlijk wel de bekendste internationale vorm van tekstbegripsmeting (Pearson & Hamm, 2005). Ook in Nederland is het een veel gebruikte toets in het basis- en voortgezet onderwijs, die ook vaak ingezet wordt om in de leesles tekstbegrip te trainen (bijv. Canton et al., 2013; Bogaerds-Hazenberg et al., 2022b). Als argument voor deze inzet in de leesles kan gelden dat het stellen van vragen bij een tekst een diepere verwerking van de tekstinhoud genereert (Rouet et al., 2001; vgl. Hendrix & Hulshof, 1994). Bovendien staat de taak een zeer gerichte bevraging van tekstkenmerken en -elementen toe. Opvallend aan de courante taakinrichting is dat een leerling vooraf doorgaans een nogal generieke, tweeledige opdracht ontvangt: 'Bestudeer de tekst goed en beantwoord daarna de vragen'.

Nemen we aan dat een vaardige lezer de tekst nog bij de vragen beschikbaar heeft én dat hij zich vanuit zijn contextmodel ten volle engageert met de taak (omdat het bijvoorbeeld een toets betreft): welk taakmodel bouwt deze lezer dan bij bovenstaande opdracht hoogstwaarschijnlijk op? We stellen nu een hypothetisch scenario op, vertrekkend vanuit de ideeën van Britt et al. (2018), dat we daarna onderzoeksmatig nader zullen onderbouwen. Een vaardige lezer ervaart vanuit de *minimal task-elaboration hypothesis* het eerste doel achter de opdracht ('de tekst goed begrijpen') hoogstwaarschijnlijk als weinig concreet; het tweede doel (het 'correct beantwoorden van de vragen') is voor hem aanmerkelijk concreter en genereert bovendien een eindproduct dat naderhand via correctie meetbaar op kwaliteit is te beoordelen. Vooraf zal hij de tekst daarom tamelijk oppervlakkig doorlezen, alleen al omdat het hem wordt verzocht maar vooral omdat het hem kan helpen de vragen uiteindelijk beter te duiden en antwoord-relevante tekstelementen beter te lokaliseren. Bij het beantwoorden van de vragen zal hij zich dan, naargelang een vraag daartoe noodzaakt, tot een nauwkeuriger en dieper bestudering van relevante tekstelementen zetten.

Dit hypothetische taakmodel van een vaardiger lezer sluit aan bij de resultaten van de meeste onderzoeken naar lees- en antwoordgedrag bij teksten met vragen. Allereerst: de initiële bestudering van de tekst. Deze draagt inderdaad tot op enige hoogte bij aan een hogere score op de vragen (Cerdán et al., 2009; Salmerón et al., 2015). Leerlingen die de tekst vooraf hebben gelezen, bouwen vooraf een globaal tekstbeeld op, waardoor ze beter de vraag interpreteren en beter de vraag-relevante tekstpassages selecteren en interpreteren (Cataldo & Oakhill, 2000; O'Reilly et al., 2018; Payne & Reader, 2006; Salmerón et al., 2015).

Dan: het antwoordproces. Vooral hier blijken vaardige lezers zich te onderscheiden van minder vaardige lezers. Het beantwoorden van een tekstafhankelijke vraag dwingt een lezer namelijk tot een veeleisende, stapsgewijze zelfregulatie. Zo dient hij allereerst de vraag te begrijpen: hij moet de mentale representatie van de vraag in lijn brengen met de eerder opgebouwde representatie van de tekstinhoud. Dit kan minder vaardige lezers aanzienlijke problemen opleveren (Brunfaut & McCray, 2015; Cerdán et al., 2019; Vidal-Abarca et al., 2010). Vervolgens dient hij zijn 'opzoekbeslissing' te reguleren: raadpleegt hij de tekst wel of niet? Minder vaardige lezers lijken vaker hun eigen begrip van de tekst en van de vraag

te overschatten dan vaardige lezers (Dunlosky & Rawson, 2005; Maki, 1998; Maki et al., 2005). Besluit een leerling de tekst te raadplegen, dan dient hij vervolgens het zoekproces te reguleren: hij moet de relevante informatie vinden en adequaat verwerken. Minder vaardige lezers besteden gemiddeld minder tijd aan relevante informatie in de tekst dan vaardige lezers (hoofdstuk 3; Bax, 2013; Cataldo & Oakhill, 2000). Ten slotte moet een lezer zijn antwoord reguleren en een antwoord formuleren of, bij een meerkeuzevraag, een antwoordalternatief selecteren (Cain & Oakhill, 2006; Traub, 1993). Het type vraag en de complexiteit van de vraag hebben op dit gehele proces een grote impact. Daardoor kan het onderscheid in antwoordgedrag tussen vaardige en minder vaardige lezers zich per vraag op zeer uiteenlopende wijze uiten: vaardige lezers lijken bijvoorbeeld niet per vraag gemiddeld minder tijd te besteden aan het antwoordproces dan minder vaardige lezers, ze kijken bijvoorbeeld korter naar de vraag en bouwen bij de tekstraadpleging meer controlestappen in (hoofdstuk 3; vgl. Bax, 2013; Brunfaut & McCray, 2015).

Ten slotte: de antwoordscore – een hypothetische verwachting. In welke mate de antwoordscore de kwaliteit van de initiële tekstbestudering of die van het antwoordproces weerspiegelt, hangt hoogstwaarschijnlijk voor belangrijk deel af van de complexiteit en ‘lokaliteit’ van de gestelde vragen. Bij eenvoudige of meer globale vragen zal een vaardiger lezer het antwoord vrijwel direct vanuit zijn reeds opgebouwde tekstrepresentatie kunnen construeren, terwijl vragen over zeer lokale tekstelementen of complexere vragen hem eerder zullen noodzaken om tamelijk omstandig de tekst te raadplegen: zijn uiteindelijke antwoord weerspiegelt dan vooral de kwaliteit van dit ‘uitgebouwde’ antwoordproces.

Deze verwachting aangaande de antwoordscore kan de nogal uiteenlopende resultaten in internationale studies op dit vlak verklaren. Er is, zoals eerder besproken, een samenhang aantoonbaar tussen vooraf-lezen en de antwoordscore (bijv. Cerdán et al., 2009) Tegelijk lijkt in een aantal studies het antwoordproces een sterk op zichzelf staand complex, waarin ‘opzoekvaardigheid naar aanleiding van de vraag’ doorslaggevend is: deze zoekbeslissingen kunnen soms duidelijk losstaan van het initiële tekstbegrip en ‘*test-management-strategies*’ betreffen (zie hoofdstuk 3; Cohen & Upton, 2006; deels: Ozuru et al., 2008; O’Reilly et al., 2018; Schroeder, 2011; Vidal-Abarca et al., 2010). Een enkele studie trekt zelfs de conclusie dat in een tekst met vragen vrijwel uitsluitend toetsvaardigheden worden aangesproken (Rupp et al., 2006). Die conclusie wordt door andere studies weer sterk gerelativeerd: mentale processen die tijdens de ‘*response decision phase*’ spelen, lijken wel degelijk samen te hangen met het tekstbegrip dat leerlingen vooraf hebben opgebouwd (Gil et al., 2015; Ozuru et al., 2008; vgl. Ardoin et al., 2019).

Maar wat als de tekst bij de vragen niet langer raadpleegbaar is? In het taakmodel van een vaardige lezer, zoals we dat hierboven beschreven en met onderzoeksresultaten onderbouwden, is de verwachting dat de tekst beschikbaar blijft bij de vragen, cruciaal: tijdens de antwoordfase weet hij immers dat hij onvolledig opgebouwd tekstbegrip altijd nog kan repareren, indien noodzakelijk. Weet hij vooraf dat hij bij de vragen de tekst niet langer beschikbaar zal hebben, dan ontstaat er een drastisch andere situatie.

We schetsen voor deze situatie weer een hypothetisch taakmodel. Wanneer vooraf bekend is dat de tekst bij de vragen niet langer raadpleegbaar is, dan rijst voor een lezer, volgens de *'parsing the context hypothesis'*, allereerst de vraag in hoeverre hij zich vanuit zijn contextmodel wil engageren met deze taak. Het doel 'de tekst goed begrijpen' is immers vaag en vergt van hem, al naargelang de lengte en complexiteit van de tekst, een mogelijk grote mentale inspanning. Vindt hij de taak voldoende relevant of hangt er voor hem veel van de taak af, dan engageert hij zich en kan hij niet anders dan tijdens de tekstbestudering aan de instructie optimaal trachten te voldoen; de vragen zal hij vervolgens moeten beantwoorden vanuit zijn optimaal opgebouwde representatie van de tekstinhoud. In hoeverre hij de vragen goed beantwoordt, is in hoge mate afhankelijk van de kwaliteit van zijn eerdere tekstbestudering.

Ook dit hypothetische taakmodel laat zich onderbouwen door bevindingen in bestaand onderzoek. Leerlingen die voorafgaande aan de leestaak weten dat ze bij de vragen de tekst niet langer beschikbaar zullen hebben, bestuderen de tekst vooraf aanmerkelijk uitgebreider, met aanmerkelijk meer herlezingen, dan leerlingen die weten dat ze de tekst tijdens de vragen nog wel beschikbaar zullen hebben (Ferrer et al., 2017; Higgs et al., 2017; vgl. Agarwal & Roediger, 2011). Bovendien lijkt in deze situatie de antwoordscore wél vooral de kwaliteit van het vooraf-lezen te weerspiegelen (Ferrer et al., 2017; Ozuru et al., 2007; Schaffner & Schiefele, 2013; Schroeder, 2011; vgl. Johnston, 1984): "When students are unable to refer back to the text, they have to access their situation model to answer the comprehension questions. Thus, the more effort readers put into organizing and integrating the text during [initial] reading, the better their comprehension performance" (Schroeder, 2011, p. 892). Of en in welke mate de context achter deze taak - *high-stakes test* vs. normale leestaak - hierop van invloed is, is bij ons weten overigens (nog) niet onderzocht.

1.4 Onderzoekshypothesen

Welke impact heeft de bestudering vooraf van een tekst op de antwoordscore van een vaardige leerling bij de tekst met vragen? In deze studie willen we de relatie tussen vooraf-lezen en vragen beantwoorden nader blootleggen door twee variabelen te manipuleren: a) de beschikbaarheid van de tekst tijdens de antwoordfase, b) de intensiteit waarmee leerlingen een tekst vooraf bestuderen. Daarbij nemen we aan dat zowel de 'tekstraadpleegbaarheid' als de intensieve initiële tekstbestudering bij de vragen de antwoordscore van leerlingen aantoonbaar zal verhogen.

We verwachten:

- a) dat leerlingen die de tekst vooraf intensief bestuderen én beschikbaar hebben bij de vragen, aanmerkelijk hoger zullen scoren dan leerlingen die de tekst intensief bestuderen maar niet langer beschikbaar hebben bij de vragen;
- b) dat naarmate leerlingen een tekst *intensiever* bestuderen, ze hoger zullen scoren bij de vragen, indien de tekst beschikbaar blijft tijdens de antwoordfase.

Deze twee onderzoekshypothesen onderzoeken we bij twee zakelijke teksten, in twee experimenten met in totaal vijf verschillende condities. Dat doen we bij vaardigere lezers in het Nederlandse voortgezet onderwijs: vwo 5-leerlingen (Expertgroep Doorlopende Leerlijnen Taal, 2008, 2009; Gubbels et al., 2019). Omdat we vooral in begrip op het situatiemodelniveau geïnteresseerd zijn, stellen we in het experiment enkel vragen die op het situatiemodel betrekking hebben en die vaardigere lezers, met of zonder de tekst beschikbaar, vanuit een adequaat situatiemodel zouden moeten kunnen beantwoorden.

2 Methode

2.1 Experimentopzet en participanten

In het eerste experiment (I) manipuleren we de tekstschikbaarheid tijdens de vraagbeantwoording: hebben leerlingen die de tekst bij de vragen nog kunnen raadplegen, een hogere score dan leerlingen die de tekst niet meer kunnen raadplegen? In dit experiment is er sprake van twee experimentele condities: conditie IA ('Raadpleegbaar'), waarin de tekst bij de vragen raadpleegbaar blijft, en conditie IB ('Niet-raadpleegbaar'), waarin de tekst bij de vragen niet raadpleegbaar blijft.

In het tweede experiment (II) manipuleren we de intensiteit van de initiële tekstbestudering: in hoeverre verhoogt een intensieve tekstbestudering vooraf de score op de vragen? Hier is er sprake van drie experimentele condities: IIA ('Intensief-lezen'), waarin de leerlingen vooraf de tekst intensief moeten bestuderen; IIB ('Lezen'), waarin de leerlingen vooraf de tekst moeten lezen maar zelf de intensiteit daarvan bepalen; en IIC ('Vragen'), waarin de leerlingen niet de tekst vooraf hoeven te bestuderen en naar eigen inzicht kunnen starten met het beantwoorden van de vragen of met het lezen van de tekst.

We nemen beide experimenten af onder vaardige lezers, vwo 5-leerlingen. Van vwo 5-leerlingen mogen we verwachten dat ze een volwaardig taakmodel hebben opgebouwd rondom de bekende leestaak tekst met vragen en dat zij aanpassingen in de taakinrichting adequaat in hun taakmodel kunnen verdisconteren. In alle condities krijgen de deelnemers hetzelfde materiaal onder ogen: twee zakelijke teksten, elk gevolgd door negen situatiemodelvragen. Wel variëren we (noodgedwongen¹) in toetsvorm: experiment I wordt afgenomen als diagnostische toets, experiment II als summatieve toets.

¹ In eerste instantie waren beide manipulaties in één experimentopzet opgenomen, als onderdeel van de summatieve tekstbegripstoetsing in het schoolvak Nederlands. Echter, deelnemende docenten Nederlands stonden zeer terughoudend tegenover een vorm van summatieve toetsing waarbij de tekst niet meer bij de vragen beschikbaar zou zijn: ze achtten de kans hoog dat onverantwoord grote verschillen in score zouden ontstaan tussen condities. Daarom besloten we tekstbegrip in experiment I ('Raadpleegbaarheid') diagnostisch te toetsen en dit in experiment II ('Intensiteit') summatief te toetsen.

Experiment I. In experiment I, waarin we de impact van tekstbeschikbaarheid op de antwoordscore onderzoeken, hebben we twee condities ingericht: IA ('Raadpleegbaar') en IB ('Niet-raadpleegbaar'). Dit experiment is afgenomen in tien vwo 5-klassen ($N = 174$) op vier voortgezet onderwijs scholen in het midden en zuiden van Nederland. Deelnemende docenten Nederlands boden leerlingen de test aan als diagnostische toets in de reguliere les Nederlands, in twee zittingen van 50 minuten. Omdat afname in klassenverband plaatsvond, werden klassen gerandomiseerd toegekend aan een van de twee experimentcondities. Leerlingen wisten vooraf niet aan welke van de twee condities hun klas deelnam. Daarnaast ontving elke klas gerandomiseerd twee versies van de toets, toetsboekjes waarin enkel de volgorde van de twee sets teksten en vragen werd omgewisseld.

Bij aanvang van de toets ontvingen de deelnemende leerlingen in beide condities de eerste tekst (lengte: 800-900 woorden), voorafgegaan door een uitgebreide instructie. Daarbij werden ze geïnstrueerd de algemene gedachtegang erin zo te bestuderen dat ze later tekstafhankelijke begripsvragen erover zouden kunnen beantwoorden zonder dat ze de tekst daarbij nog raadpleegbaar hadden: de tekst zou na bestudering worden ingenomen, waarna de vragen zouden worden verstrekt. Onderstrepingen en aantekeningen in de tekst waren toegestaan, alsmede aantekeningen op een apart blad, die echter zouden worden ingenomen met de teksten. Voor deze tekstbestudering hadden leerlingen per tekst een behoorlijke hoeveelheid tijd beschikbaar: 15 minuten. Leerlingen konden de tekst niet eerder inleveren, noch konden ze de vragen eerder opvragen. Met deze instructie hoopten we de leerlingen in experiment I te dwingen tot een zeer intensieve bestudering van de tekst. Immers, lezers lezen vooraf een tekst doorgaans aanmerkelijk langer en met meer aandacht voor de hoofdzaken wanneer ze weten dat ze een tekst niet beschikbaar zullen hebben bij de vragen (Ferrer et al., 2017; Schroeder, 2011).

In conditie IA ('Raadpleegbaar') werden de leerlingen met deze instructie echter misleid: bij het overhandigen van de vragen werd de tekst uiteindelijk *niet* ingenomen. De surveillant nam dus na vijftien minuten de tekst niet in en overhandigde enkel de vragen, waarbij deze aangaf dat ze tekst wel degelijk konden raadplegen. Deze procedure werd in de tweede les herhaald, met de andere set tekst en vragen.²

In conditie IB ('Niet-raadpleegbaar') nam de surveillant na vijftien minuten de tekst wél in en overhandigde de vragen. Hier beschikten de leerlingen dus bij het beantwoorden van de vragen, conform instructie, niet langer over de tekst. Deze procedure werd in de tweede les herhaald, met de andere set tekst en vragen.

Leerlingen dienden voorafgaand aan het beantwoorden van de vragen als eerste op het antwoordblad aan te geven hoeveel minuten ze de tekst vooraf hadden bestudeerd. De surveillant controleerde of ze dit daadwerkelijk noteerden.

² Bij de onderzoekscondities IA ('Raadpleegbaar') en IIA ('Intensief-lezen') werden de leerlingen misleid: de leesteksten zouden bij de vragen worden ingenomen, wat uiteindelijk niet gebeurde. Het ligt voor de hand aan te nemen dat leerlingen bij de tweede test deze misleiding met een korreltje niet zullen hebben genomen. In de analyse is daarom erop gecontroleerd of een mogelijk verminderde effectiviteit van deze misleiding haar sporen heeft nagelaten in de resultaten in het experiment. Dit bleek echter niet aantoonbaar.

Experiment II. In experiment II, waarin we de intensiteit van de tekstbestudering vooraf manipuleren, hebben we drie condities ingericht: de condities IIA ('Intensief-lezen'), IIB ('Lezen') en IIC ('Vragen'). Dit experiment is afgenomen in zestien vwo 5-klassen ($N = 398$) op vier voortgezet onderwijscholen in het midden en zuiden van Nederland. Hier maakte de toets onderdeel uit van de summatieve toetsing in de toetsweek, waarbij deze werd afgenomen in één centrale zitting van 100 minuten. Docenten van de deelnemende scholen surveilleerden bij de toets. Omdat afname steeds plaatsvond op een centraal toetsmoment, konden leerlingen (en dus niet klassen, zoals bij experiment I) binnen scholen gerandomiseerd worden toegekend aan een van de drie experimentcondities. Leerlingen wisten vooraf niet aan welke conditie ze deelnamen, noch dat er sprake was van onderling afwijkende afnamecondities. In elke afzonderlijke testruimte had telkens slechts één conditie zitting. Daarnaast ontving elke groep gerandomiseerd twee versies van de toets, waarin enkel de volgorde van de sets afweek.

Conditie IIA ('Intensief-lezen') is in opzet, instructie en procedure gelijk aan conditie IA ('Raadpleegbaar') uit experiment I: we herhalen daarom hier enkel de hoofdzaken. In conditie IIA ('Intensief-lezen') ontvingen de leerlingen allereerst een tekst met daarbij de instructie de algemene gedachtegang erin zodanig te bestuderen dat ze straks begripsvragen erover zouden kunnen beantwoorden zonder de tekst te kunnen raadplegen: de tekst zou dan worden ingenomen. Daarvoor hadden ze per tekst 15 minuten beschikbaar. Uiteindelijk bleken leerlingen hiermee te zijn misleid, want bij het overhandigen van de vragen werd de tekst uiteindelijk niet ingenomen. Zo hoopten we een intensieve vooraf-bestudering van de tekst bij leerlingen af te dwingen. Na vijftig minuten herhaalde zich deze procedure.

In conditie IIB ('Lezen') wilden we leerlingen ertoe aanzetten dat ze minimaal de hele tekst gelezen hadden voordat ze starten met het beantwoorden van de vragen. In deze conditie ontvingen de leerlingen eerst een tekst, met de instructie de algemene gedachtegang zo te bestuderen dat ze straks begripsvragen erover zouden kunnen beantwoorden. Die begripsvragen zouden na 15 minuten worden uitgereikt; de leerlingen vernamen hier dus expliciet dat de tekst bij de vragen raadpleegbaar zou blijven. Desgewenst kon een leerling bij de surveillant na lezing wel eerder om de vragen verzoeken. Doordat deze leerlingen wisten dat de tekst raadpleegbaar bleef bij de vragen en ze de optie hadden de vragen eerder op te vragen, verwachtten we dat deze leerlingen vooraf de tekst zouden bestuderen maar daarbij een minder rijk gespecificeerd begrip van de tekst zouden opbouwen dan de leerlingen in de vorige conditie.

In conditie IIC ('Vragen') verliep de procedure grotendeels zoals leerlingen doorgaans bij leestoetsen in het schoolvak Nederlands gewend zijn. Hier ontvingen de leerlingen bij de start van het experiment een vragenboekje, waarin achterin de tekst als bijlage was opgenomen. De instructie op de toets was in deze conditie minimaal: leerlingen werd slechts geïnstrueerd de begripsvragen te beantwoorden, en daarbij werd vermeld dat ze niet noodzakelijkerwijs de tekst vooraf hoefden te bestuderen. Wanneer leerlingen klaar waren, konden ze het materiaal

inleveren; na vijftig minuten haalde de surveillant al het materiaal op. Daarna overhandigde hij de tweede set tekst en vragen, met dezelfde instructie als bij de eerste set. Doordat leerlingen in deze conditie niet werden aangespoord tot initiële tekstbestudering, verwachtten we dat ze de tekst vooraf enkel globaal of gedeeltelijk zouden bestuderen – of mogelijk zelfs direct zouden starten met de vragen – en zo vooraf een minder diep tekstbegrip zouden hebben opgebouwd dan de leerlingen in de twee voorgaande condities.

2.2 Voorbereiding op deelname aan experiment

Vooraf hadden deelnemende docenten Nederlands aangegeven dat hun vwo 5-leerlingen niet vertrouwd waren met een vorm van tekstbegripstoetsing waarbij de tekst niet langer raadpleegbaar was bij de vraagbeantwoording. We wilden niet dat leerlingen uit onervarenheid met deze toetsingsvorm tijdens het experiment zouden worden verrast en mogelijk een inadequaat taakmodel zouden opstellen. Daarom bespraken deelnemende docenten enkele weken voorafgaande aan het experiment aan de hand van een PowerPoint met hun klassen welke implicaties (niet-)tekstraadpleegbaarheid heeft voor de initiële tekstbestudering en voor de vraagbeantwoording.

Daarbij ontvingen de leerlingen ook een korte introductie op het experiment. De docent gaf aan dat de toets onderdeel uitmaakte van een landelijke peiling van het tekstbegrip van vwo-leerlingen, naar aanleiding van recente berichten over een terugval in tekstbegrip in Nederland, onder anderen bij vwo-leerlingen (vgl. Gubbels et al., 2019).

2.3 Constructie van het materiaal

Omdat we nadrukkelijk het bestaande context- en taakmodel van vwo 5-leerlingen bij tekstbegripstoetsen wilden aanspreken, werd bij het samenstellen van het materiaal zoveel mogelijk aangesloten op de reguliere toetsing van tekstbegrip in het voortgezet onderwijs. In havo- en vwo-bovenbouw wordt vaak geoefend met onderdelen uit de Centraal Schriftelijke Examens (CSE's) Nederlands havo en vwo (Canton et al., 2013). Als materiaal werden daarom twee zakelijke opiniërende 'teksten met vragen' geselecteerd uit CSE's havo en vwo.

Bij de selectie van de CSE-onderdelen golden drie criteria. 1) Er werd allereerst op gelet dat de teksten een heldere tekststructuur kenden: een vaardige lezer moest al bestuderend in staat zijn een duidelijke, samenhangende gedachtegang in de tekst aan te wijzen. 2) Daarnaast moesten de CSE-vragen vooral een appel doen op het situatiemodel van een lezer. We selecteerden daarbij de vragen die als *elaboration questions* of *bridging inference questions* konden worden beschouwd, typerende situatiemodelvragen (Kintsch, 1998). *Elaboration questions* vragen om de verbinding van een tekstpassage met de achtergrondkennis van een lezer; *bridging inference questions* vereisen dat een lezer twee of meer tekstpassages inhoudelijk verbindt (vgl. Kintsch, 1998; Land, 2009; O'Reilly & McNamara, 2007). 3) Ten

slotte zochten we CSE-onderdelen die een relatief hoge betrouwbaarheid kenden, waarbij we gebruik konden maken van interne Cito-rapportages. Op grond hiervan selecteerden we uiteindelijk één onderdeel uit het CSE vwo 2019-1 ('De Kunst en de regels') en één onderdeel uit het CSE havo 2017-2 ('De wilskrachtige is een gewoontedier').

Na selectie werden de teksten in deze CSE-onderdelen enigszins ingekort (max. ± 900 woorden) om intensieve bestudering binnen een relatief kort tijdsbestek te vergemakkelijken en de noodzakelijke mentale inspanning van deelnemers niet te overvragen, vooral van de groep deelnemers die de vragen zonder de tekst zouden moeten maken. In Bijlage 2 zijn beide sets teksten en vragen opgenomen.

Deze twee CSE-onderdelen kenden volgens Citogegevens een betrouwbaarheid van $GLB = .46$ en $GLB = .41$, relatief lage waarden, waarin we moeten verdisconten dat elk van beide onderdelen slechts een beperkt aantal items bevat (resp. 9 en 10 items); binnen de context van het CSE Nederlands havo/vwo presteerden deze twee CSE-onderdelen qua betrouwbaarheid evenwel bovengemiddeld.³

Vragen werden zoveel mogelijk overgenomen uit de betreffende CSE-onderdelen. Vragen die vooral een appel deden op de tekstbasis werden daarbij verwijderd, evenals analytische vragen over alineafuncties, drogredenen en argumentatieschema's. Slechts vragen op het situatiemodelniveau, vragen die (onderdelen van) de centrale gedachtegang van de tekst bevroegen, werden gehandhaafd. Het betrof zowel open als gesloten vragen. Op basis van inhoudelijk commentaar van de vakvereniging van docenten Nederlands op een van beide CSE-onderdelen (LTN, 2019a), werd een klein aantal onvolkomenheden in de originele vraagstelling aangescherpt. Daarnaast werden in beide onderdelen enkele vragen zo aangepast dat leerlingen ze konden beantwoorden zónder dat ze de tekst nog konden raadplegen. Een tweetal vragen ten slotte werd door de onderzoekers nieuw geconstrueerd. Een screeningspanel, bestaande uit drie gepromoveerde onderzoekers in taalbeheersing en drie geroutineerde bovenbouwdocenten Nederlands, screende daarop al het materiaal uitvoerig; daarnaast stelde het panel het correctiemodel vast.

De deelnemende docenten Nederlands corrigeerden de toets op basis van het vastgestelde correctiemodel. Een van de onderzoekers corrigeerde daarnaast afzonderlijk al het materiaal; afwijkingen in score werden in overleg in overeenstemming gebracht.

³ Ter vergelijking, gebaseerd op de gegevens van de laatste drie jaar: CSE-onderdelen met een beperkt aantal vragen zoals in deze twee sets kennen doorgaans een nogal lage betrouwbaarheid ($GLB < .40$), terwijl een CSE havo/vwo in zijn geheel, met gemiddeld drie tot vijf teksten, met gemiddeld dertig tot veertig vragen, een GLB van circa $.70$ kent.

In Nederland is het gebruik van GLB (*Greatest Lower Bound*) als maat voor betrouwbaarheid overigens nog relatief onbekend, terwijl deze maat een accuratere schatting hiervan geeft dan de traditionele maat, Cronbach's α , die altijd een ondergrensschatting is van de werkelijke betrouwbaarheid. Zie hiervoor Sijtsma (2009) en Trizano-Hermosilla en Alvarado (2016).

2.4 Data-analyse

Controle manipulatie. Om te achterhalen in hoeverre leerlingen vooraf de teksten inderdaad in wisselende intensiteit per conditie bestudeerden, ontwikkelden we een aantal controle-instrumenten.

Ten eerste vroegen we leerlingen in alle condities – behalve de leerlingen in conditie IIC ('Vragen'), die de tekst niet hoefden te bestuderen – voorafgaand aan het maken van de vragen de tijdsduur aan te geven van hun tekstbestudering. We gaan ervan uit dat, als leerlingen in de ene conditie gemiddeld langer een tekst bestudeerden dan de leerlingen in een andere conditie, ze daarmee blijken van meer cognitieve verwerkingstijd en dus intensiever de tekst hebben bestudeerd.

Ten tweede scoorden we in de tekstboekjes van alle condities of a) leerlingen tekstelementen onderstreepten of *highlighten*, en daarnaast b) of leerlingen bij de tekst aantekeningen maakten, al dan niet op een apart blad. Daarbij gaan we ervan uit dat leerlingen intensiever de tekst hebben bestudeerd naarmate ze vaker onderstrepen en/of aantekeningen maken. Voor het scoren van de onderstrepingen en aantekeningen op de leesteksten maakten we gebruik van een tweepuntsschaal. Hierbij gold '0': leerlingen onderstrepen niet / maken geen aantekeningen; '1': leerlingen onderstrepen wel / leerlingen maken geen aantekeningen.

Analyse. Voor elke leerling is allereerst de totaalscore berekend voor de eerste set tekst en vragen én voor de tweede set tekst en vragen afzonderlijk. Voor de analyse van de verschillen in leestijd per set én de gemiddelde antwoordscore tussen de condities wordt vervolgens gebruikgemaakt van multiniveau-modellen, waarbij we als *fixed effects* de conditie en set 'tekst met vragen' 1 of 2 meenamen, als hoofdeffect en interactie-effect. Daarbij zijn scores steeds genest binnen leerlingen en klassen.

Om te bezien of er verschillen bestaan in de hoeveelheid onderstrepingen en aantekeningen tussen de condities, stelden we hierbij vooraf een binomiaal model op, waarbij we condities en 'eerste/tweede set' als hoofdeffecten en interactie-effect meenemen en we een variantie schatten tussen klassen.

3 Resultaten

3.1 Algemeen

Experiment I. Hierin verliep de afname grotendeels zoals beoogd. De leerlingen ($N = 174$) zijn redelijk evenwichtig verspreid over de twee condities (conditie IA 'Raadpleegbaar': $n = 77$; conditie IB 'Niet-raadpleegbaar': $n = 97$), waarbij jongens en meisjes niet afwijkend verspreid waren over de condities ($\chi^2(1) = 0.53$; $p = .47$). Deelnemende docenten Nederlands gaven aan dat leerlingen doorgaans voldoende geconcentreerd en gemotiveerd waren.

Op twee van de vier scholen verliep de afname zonder problemen, al was er door de Covid-epidemie wat meer lesabsentie dan gebruikelijk. Op de twee andere

scholen werd afname van het experiment lange tijd verhinderd door de tweede lockdown van het Nederlandse voortgezet onderwijs, vanaf december 2020. Pas na de hervatting van het reguliere onderwijs in juni 2021, kon het experiment worden afgenomen, waarbij de ene school het materiaal enkel in gehalveerde klassen kon afnemen en de andere school met veel absentie te kampen had.

De twee testen tezamen bleken uiteindelijk in hoge mate betrouwbaar te differentiëren tussen de twee condities ($GLB = .86$).

Experiment II. Hierin verliep de afname zoals beoogd. De leerlingen ($N = 398$) zijn redelijk evenwichtig verspreid over de drie condities (conditie IIA 'Intensief-lezen': $n = 139$; conditie IIB 'Lezen': $n = 113$; conditie IIC 'Vragen': $n = 146$), waarbij jongens en meisjes niet afwijkend verspreid waren over de condities ($\chi^2(2) = 2.94$; $p = .23$). Surveillanten en deelnemende docenten Nederlands gaven aan dat leerlingen gemotiveerd en geconcentreerd deelnamen aan de toets. Wat conditie IIC ('Vragen') betreft, merkten surveillanten op dat het merendeel van de leerlingen na het uitdelen van de toetsboekjes niet startte met de vragen, maar met het lezen van de tekst.

De twee testen tezamen bleken voldoende betrouwbaar te differentiëren tussen de drie condities ($GLB = .64$).

3.2 Controle manipulatie

Vooraf-leestijden: experiment I. Hier vinden we voor de gemiddelde vooraf-leestijd allereerst een hoofdeffect van conditie ($F(1, 144.23) = 6.21$; $p = .01$). Leerlingen in conditie IA ('Raadpleegbaar') besteden gemiddeld 12.20 minuut ($se = .52$) aan een tekst terwijl de leerlingen in conditie IB ('Niet-raadpleegbaar') gemiddeld 12.88 minuut ($se = .50$) aan het vooraf-lezen van een tekst besteden.

We kunnen niet aantonen dat de ene tekst vooraf langer of korter wordt gelezen dan de andere ($F(1, 154.89) = .28$; $p = .60$), noch vinden we een interactie-effect tussen tekst en conditie ($F(1, 237.97) = .88$; $p = .35$): we kunnen dus niet aantonen dat de ene tekst korter of langer wordt gelezen dan de andere tekst, afhankelijk van de conditie.

Vooraf-leestijden: experiment II. Hier vinden we dat leerlingen in conditie IIA ('Intensief-lezen') beide teksten vooraf duidelijk anders lezen dan leerlingen in conditie IIB ('Lezen') ($F(1, 242.95) = 50.92$; $p < .001$). De leerlingen in conditie IIA ('Intensief-lezen') besteden vooraf gemiddeld 14.31 minuut ($se = .28$, voor tekst 2) aan het vooraf-lezen terwijl leerlingen in conditie IIB ('Lezen') gemiddeld 11.86 ($se = .31$) eraan besteden.

Er blijkt ook sprake van een verschil tussen teksten ($F(1, 241.95) = 5.91$; $p = .02$). Leerlingen blijken aan tekst 1, de vwo-tekst, gemiddeld .49 minuut ($se = .19$) meer tijd te besteden dan aan tekst 2, de havotekst.

Er is evenwel geen sprake van een interactie-effect tussen tekst en conditie ($F(1, 241.95) = .27$; $p = .61$).

Onderstrepingen en aantekeningen: experiment I. In 40 procent van de teksten zijn elementen *onderstreept*. Dit onderstrepen verschilt niet aantoonbaar tussen condities ($F(1, 313) = 3.12; p = .08$), noch tussen tekst 1 en tekst 2 ($F(1, 313) = 2.17; p = .14$). We kunnen evenmin een interactie-effect tussen tekst en conditie aantonen ($F(1, 313) = 3.68; p = .06$).

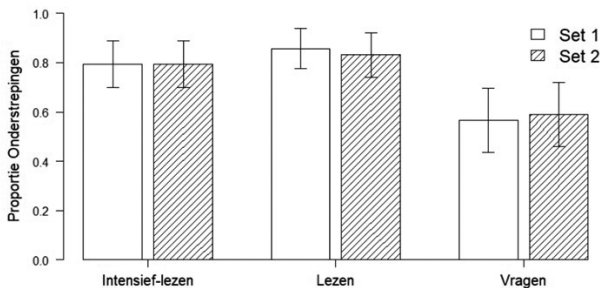
In 13 procent van de teksten zijn *aantekeningen gemaakt*. Dit aantekeningen maken verschilt niet aantoonbaar tussen condities ($F(1, 313) < .001; p > .99$), noch tussen tekst 1 of tekst 2 ($F(1, 313) = .68; p = .41$). We kunnen ook geen interactie-effect tussen tekst en conditie aantonen ($F(1, 313) = .82; p = .37$).

In dit experiment vinden we dus, zoals verwacht, geen verschillen tussen condities op de controlevariabelen.

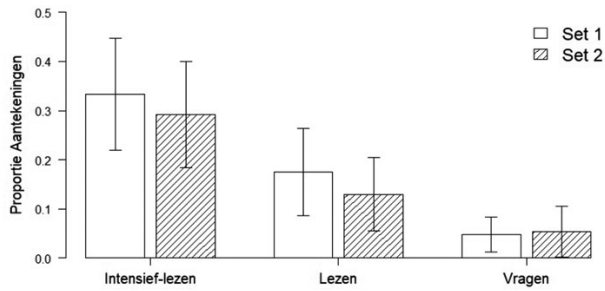
Onderstrepingen en aantekeningen: experiment II. Per conditie blijken leerlingen af te wijken in de gemiddelde hoeveelheid *onderstrepingen* ($F(2, 789) = 23.48; p < .001$). In 79.4 procent ($se = .04$, dus 4 procent) van de teksten van leerlingen in conditie IIA ('Intensief-lezen') zijn zaken *onderstreept* en in de teksten van leerlingen in conditie IIB ('Lezen') is dat 84.4 procent ($se = .04$). In de teksten van leerlingen in conditie IIC ('Vragen') is dat 57.8 procent ($se = .06$). De leerlingen in conditie IIA en IIB onderstrepen daarmee significant vaker ($p < .001$) dan de leerlingen in conditie IIC. Leerlingen in conditie IIA onderstrepen echter niet aantoonbaar meer of minder dan de leerlingen in conditie IIB ($p = .88$).

We kunnen niet aantonen dat leerlingen in tekst 1 minder of meer onderstrepen dan in tekst 2 ($F(1, 789) = .03; p = .85$), noch kunnen we een interactie-effect tussen tekst en conditie aantonen ($F(2, 789) = .22; p = .80$). In Figuur 4.1 zijn de geschatte gemiddelden voor de onderstrepingen per conditie per tekst weergegeven.

Wat *aantekeningen maken* betreft, verschillen condities significant ($F(2, 789) = 30.76; p < .001$). Van de teksten van leerlingen in conditie IIA ('Intensief-lezen') bevat 31.2 procent aantekeningen ($se = .05$). Voor de teksten van leerlingen in conditie IIB ('Lezen') is dat 15.1 procent ($se = .03$), en voor de teksten van leerlingen in conditie IIC ('Vragen') 5.1 procent ($se = .02$). Daarmee onderstrepen de leerlingen in conditie IIA beduidend vaker dan de leerlingen in conditie IIB ($p < .01$) en vaker dan de leerlingen in conditie IIC ($p < .001$). Leerlingen in conditie IIB onderstrepen



Figuur 4.1 Proportie onderstrepingen bij een tekst, voor set 1 en 2, per conditie, voor experiment II ('Intensiteit')



Figuur 4.2 Proportie aantekeningen maken bij een tekst, voor set 1 en 2, per conditie, voor experiment II ('Intensiteit')

daarmee ook vaker dan de leerlingen in conditie IIC ($p = .03$). We kunnen dan ook concluderen dat de manipulatie in experiment II geleid heeft tot een andere taak-aanpak tussen de condities.

Leerlingen blijken in tekst 1 niet aantoonbaar meer of minder aantekeningen te maken dan in tekst 2 ($F(1, 789) = .40; p = .53$). Er is evenmin sprake van een interactie-effect tussen tekst en conditie ($F(2, 789) = .29; p = .75$). In Figuur 4.2 zijn de geschatte gemiddelden voor het aantekeningen maken per conditie per tekst weergegeven.

3.3 Antwoordscore

Experiment I. Kijken we vervolgens naar de gemiddelde antwoordscore, dan vinden we allereerst een duidelijk verschil tussen condities ($F(1, 110.62) = 22.44; p < .001$). In Tabel 4.1 zijn de schattingen voor de antwoordscore weergegeven voor experiment 1. We zien dan dat de leerlingen in conditie IB ('Niet-raadpleegbaar') per set aanmerkelijk lager scoren dan de leerlingen in conditie IA ('Raadpleegbaar'): leer-

Tabel 4.1 Parameterschattingen voor de antwoordscore bij experiment I ('Raadpleegbaarheid')

Parameter	coëfficiënt	(se)
<i>Conditie</i>		
conditie IA 'Raadpleegbaar': set tekst en vragen 2	9.01	(.35)
conditie IB 'Niet-raadpleegbaar': set tekst en vragen 2	6.54	(.33)
<i>Set</i>		
Δ set tekst en vragen 1	-.10 ^a	(.42)
<i>Conditie * Set</i>		
Δ conditie IA 'Raadpleegbaar': set tekst en vragen 1	-1.41	(.65)
<i>Varianties</i>		
S ² klas	.37	(.33)
<i>(Co)varianties tussen leerlingen</i>		
	S ² set 1	S ² set 2
S ² set 1	13.55 (1.56)	.37 ^b
S ² set 2	5.43 (.62)	2.93 (.76)

^a niet-significant; ^b correlatie tussen set 1 en set 2

lingen in conditie IA ('Raadpleegbaar') behalen een gemiddelde score van 9.01 ($se = .35$, voor de tweede set tekst en vragen), terwijl leerlingen in conditie IB ('Niet-raadpleegbaar') een gemiddelde score van 6.54 ($se = .32$) behalen.

We vinden eveneens een verschil in score tussen de eerste set tekst en vragen en de tweede set tekst en vragen ($F(1, 164.88) = 7.60$; $p < .01$), zoals er ook sprake blijkt van een interactie-effect tussen de sets en condities ($F(1, 238.82) = 4.75$; $p = .03$). Leerlingen in conditie IA ('Raadpleegbaar') blijken bij de eerste set tekst en vragen gemiddeld -1.41 ($se = .65$) lager te scoren dan bij de tweede set tekst en vragen. In Figuur 4.3 zijn de gemiddelde scores voor de beide sets in de twee condities weergegeven.

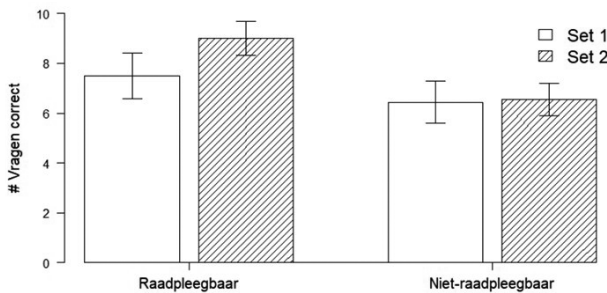
Leerlingen in conditie IB ('Niet-raadpleegbaar' scoren dus gemiddeld steeds lager dan de leerlingen in conditie IA ('Raadpleegbaar'), al is dit onderscheid in score tussen de condities groter voor de tweede set (de havo-set) dan voor de eerste set (de vwo-set).

Onderaan in Tabel 4.1 zijn de varianties weergegeven. We zien dan dat de variantie tussen klassen zeer gering is ($.37$, $se = .33$) en dat de variantie bij de eerste set, het onderdeel uit het vwo-examen, zeer hoog is (13.55 , $se = 1.56$). De variantie bij de tweede set, het onderdeel uit het havo-examen, is daarmee vergeleken duidelijk lager (5.43 , $se = .62$). Dit duidt erop dat de vwo-set meer verschillen tussen leerlingen laat zien dan de havo-set (en dus waarschijnlijk aanmerkelijk pittiger is).

De covariantie tussen beide sets is 2.93 ($se = .76$), waarbij de correlatie tussen beide $.37$ is.

Experiment II. Vonden we eerder duidelijke verschillen in gemiddelde antwoordscore tussen de condities, bij het tweede experiment kunnen we hierin geen verschillen aantonen ($F(2, 395.76) = .55$; $p = .58$). In Tabel 4.2 zijn de schattingen voor de gemiddelde antwoordscore weergegeven: ook niet-significante parameterschattingen zijn hierin opgenomen. Daarin is te zien dat de leerlingen in conditie IIC ('Vragen') in absolute zin hoger scoren dan de leerlingen in de twee andere condities - zie Tabel 4.2 -, evenwel, dit onderscheid tussen condities is niet significant.

Wel vinden we een verschil in score tussen de eerste en de tweede set tekst en vragen ($F(1, 397.47) = 5.11$; $p = .02$): leerlingen scoren bij de eerste set over het algemeen $.39$ ($se = .16$) hoger dan bij de tweede set.



Figuur 4.3 Gemiddelde antwoordscore voor set 1 en 2, per conditie, voor experiment I ('Raadpleegbaarheid')

Tabel 4.2 Parameterschattingen voor de antwoordscore bij experiment II ('Intensiteit')*

Parameter	coëfficiënt	(se)
<i>Conditie</i>		
conditie IIA ('Intensief-lezen'): set tekst en vragen 2	9.86	(.20)
conditie IIB ('Lezen'): set tekst en vragen 2	10.15	(.21)
conditie IIC ('Vragen'): set tekst en vragen 2	10.29	(.19)
<i>Set</i>		
Δ set tekst en vragen 1	.39	(.17)
<i>Variaties</i>		
S ² klas	.15	(.11)
(Co)varianties tussen leerlingen	S ² set 1	S ² set 2
S ² set 1	9.91 (.71)	.24 ^a
S ² set 2	4.36 (.32)	1.57 (.34)

^a correlatie tussen set 1 en set 2

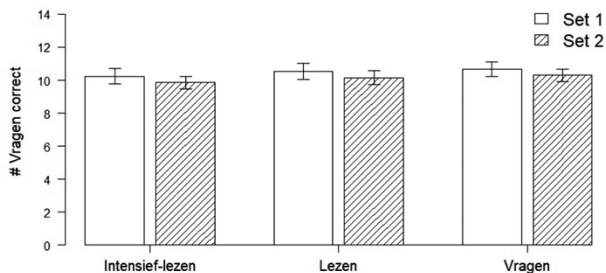
* Het interactie-effect Conditie*Set is weggelaten in de hier gerapporteerde schattingen.

Een interactie-effect tussen de sets en condities is evenwel niet aantoonbaar ($F(2, 397.48) = 2.05; p = .13$); in Tabel 4.2 is deze voor het berekenen van de schattingen daarom achterwege gelaten. In Figuur 4.4 zijn de gemiddelde scores voor de beide sets voor de drie condities weergegeven.

Onderaan in Tabel 4.2 zijn weer de varianties weergegeven. We zien dat de variantie tussen klassen erg beperkt is (.15, $se = .11$). De variantie in leerlingsscores bij de eerste set tekst en vragen (het vwo-examen) is wederom erg hoog (9.85, $se = .71$), terwijl de variantie bij de tweede set (het havo-examen) aanzienlijk lager is (4.35, $se = .31$): de pittige vwo-set lijkt weer tot meer verschillen tussen leerlingen te leiden dan de wat eenvoudiger havo-set. De covariantie tussen beide sets is relatief beperkt (1.59, $se = .34$), waardoor de correlatie tussen beide tests .24 bedraagt.

3.4 Samenvatting resultaten

Experiment I: raadpleegbaarheid. In de eerste variant, waarin het experiment als diagnostische toets werd afgenomen, kunnen we weinig tot geen verschillen in



Figuur 4.4 Gemiddelde antwoordscore voor set 1 en 2, per conditie, voor experiment II ('Intensiteit')

lees- en antwoordgedrag aantonen tussen de twee condities. Leerlingen in conditie IA ('Raadpleegbaar') onderstrepen niet meer of minder dan de leerlingen in conditie IB ('Niet-raadpleegbaar'), noch maken ze meer of minder aantekeningen. De leerlingen in conditie IA besteden wat minder tijd aan het lezen van de tekst dan de leerlingen in conditie IB.

Wat de antwoordscore betreft, vinden we een aanmerkelijk verschil tussen de condities. De leerlingen in conditie IA ('Raadpleegbaar') scoren beduidend hoger dan de leerlingen in conditie IB ('Niet-raadpleegbaar').

Experiment II: intensiteit. In het tweede experiment, dat als summatieve toets werd afgenomen, zien we allereerst duidelijke verschillen in leesgedrag tussen de drie condities. Voor de leerlingen in conditie IIA ('Intensief-lezen') geldt dat ze een tekst langer lezen en vaker aantekeningen maken bij een tekst dan de leerlingen in conditie IIB ('Lezen'); bovendien onderstrepen de leerlingen in conditie IIA ('Intensief-lezen') vaker tekstelementen dan de leerlingen in conditie IIC ('Vragen'), zij het niet vaker dan de leerlingen in conditie IIB ('Lezen').

De leerlingen in conditie IIB ('Lezen') maken weer vaker aantekeningen en onderstrepen vaker tekstelementen dan de leerlingen in conditie IIC ('Vragen'). De leerlingen lijken dus per conditie, zoals beoogd, in wisselende mate van intensiteit de tekst hebben bestudeerd. *Grosso modo* kunnen we dus stellen: naarmate in de conditie intensiever vooraf-leesgedrag werd afgedwongen, gingen leerlingen meer onderstrepen en aantekeningen maken.

We konden in dit experiment echter – tegen de verwachting in – geen verschil in antwoordscore aantonen tussen de drie condities.

Beide experimenten. Conditie IA ('Raadpleegbaar') en conditie IIA ('Intensief-lezen') wijken inhoudelijk niet af in materiaal en instructie, maar wel in hun taakcontext: diagnostische vs summatieve toets. Leerlingen in conditie IIA ('Intensief lezen'), die de summatieve toets maakten, scoren aanmerkelijk hoger op de vragen dan de leerlingen in conditie IA ('Raadpleegbaar'), die de diagnostische toets maakten, waarschijnlijk doordat deze eerste groep meer geconcentreerd en gemotiveerd deelnam.

Daarnaast: de variantie bij de eerste set (het vwo-examenonderdeel) is, zeker in vergelijking met de variantie bij de tweede set (het havo-examenonderdeel), bij beide experimenten zeer hoog. Deze eerste set lijkt al met al beduidend moeilijker te zijn geweest voor leerlingen. Deze moeilijkheid lijkt ook een mogelijke verklaring voor het feit dat in experiment I ('Raadpleegbaarheid') leerlingen bij set 1 lager scoren dan bij set 2, terwijl ze in experiment II ('Intensiteit') op set 1 juist hoger scoren dan op set 2: door de cognitieve inspanning die de vwo-set vergt, zetten leerlingen zich in de diagnostische context specifiek voor deze set waarschijnlijk minder in dan in de summatieve context.

4 Discussie en conclusie

In deze studie wilden we de relatie tussen het voorafbestuderen en de antwoordscore bij teksten met vragen nader blootleggen. Dat deden we door in twee experimenten twee variabelen te manipuleren: enerzijds de intensiteit van tekstbestudering wanneer vaardige lezers vooraf een tekst lezen, en anderzijds de mogelijkheid om de tekst te raadplegen wanneer vaardige lezers vragen beantwoorden.

Onze eerste hypothese luidde: 'Leerlingen die de tekst vooraf intensief bestuderen én beschikbaar hebben bij de vragen, zullen aanmerkelijk hoger scoren dan leerlingen die de tekst intensief bestuderen en niet langer beschikbaar hebben bij de vragen.' Dit konden we zeer duidelijk aantonen. De verwachte, doorslaggevende rol die tekstraadpleging tijdens het antwoordproces heeft voor de antwoordscore, vinden we hier bevestigd. Doordat leerlingen via tekstraadpleging hun voor de vraag inadequate of ontbrekende tekstbegrip kunnen repareren en hun conceptantwoorden kunnen controleren, wordt de antwoordscore in sterke mate beïnvloed door de tekstraadpleegbaarheid tijdens de vraagbeantwoording.

Onze tweede hypothese luidde: 'Naarmate leerlingen een tekst *intensiever* bestuderen, zullen ze hoger scoren bij de vragen, indien de tekst beschikbaar blijft tijdens de antwoordfase'. Ondanks duidelijke, verwachte verschillen in leesgedrag tussen de drie condities in experiment II, konden we dit niet aantonen. Als een intensieve tekstbestudering de antwoordscore beïnvloedt, dan ontbreekt daarvoor in deze studie elke aanwijzing.

Dat we deze tweede hypothese niet hebben kunnen aantonen, valt des te meer op omdat we in experiment II daartoe een optimale context creëerden. De deelnemers waren vaardige, op de leestaak voorbereide lezers die, naar surveillanten aangaven, geconcentreerd werkten aan een toets met een relatief zware weging voor het vak Nederlands. Deze lezers zullen, zogezegd, vaak gewerkt hebben vanuit een krachtig, adequaat context- en taakmodel. Daar komt nog bij dat ze uitsluitend situatiemodelvragen kregen voorgelegd, verbonden aan de centrale gedachtegang van de tekst. Als leerlingen vooraf zich deze centrale gedachtegang hebben eigengemaakt, dan zouden ze daarna juist deze vragen gemakkelijker (en dus vaker goed) moeten kunnen beantwoorden. Echter, zelfs onder deze vrijwel optimale omstandigheden konden we niet aantonen dat een intensieve tekstbestudering tot een hogere antwoordscore leidt. Kennelijk voegt een *intensieve* tekstbestudering weinig tot niets aantoonbaar toe aan de uiteindelijke kwaliteit van het antwoord. Daarmee lijkt de eerdere conclusie uit experiment I andermaal te worden bevestigd: de tekstraadpleging tijdens de vragen geeft vooral de doorslag voor de antwoordscore – en niet zozeer een diepgaande tekstbestudering vooraf.

Het hypothetische taakmodel van een vaardiger lezer bij de taak tekst met vragen, zoals we dat in paragraaf 1.3 onderbouwden, kunnen we dus met deze bevindingen grotendeels bevestigen. Wanneer de tekst bij de vragen beschikbaar blijft, beïnvloedt vooral het antwoordproces de antwoordscore. Het vooraf-leesproces is daarbij instrumenteel: het staat lezers toe bij de beantwoording de vraagstelling beter te duiden en beter en sneller antwoord-relevante tekstdelen te lokaliseren

(vgl. hoofdstuk 3; Bax, 2013; Brunfaut & McCray, 2015). Aan dat hypothetische taakmodel kunnen we nu toevoegen dat een intensief vooraf-leesproces bij complexe situatiemodelvragen hoogstwaarschijnlijk geen positief effect heeft op het uiteindelijke antwoord. Wat leerlingen bij het vooraf-lezen nog hebben nagelaten te doorgronden, kunnen ze afdoende repareren tijdens het antwoordproces als de tekst dan beschikbaar blijft. Het volstaat om de tekst vooraf met enige aandacht door te nemen, want begripsconstructie tijdens het antwoordproces is het enige waar het bij deze taak om draait.

De impact van de taakcontext op de prestaties van leerlingen blijkt daarbij opvallend groot. Terwijl de onderzoekscondities IA ('Raadpleegbaar') en IIA ('Intensief-lezen') zich enkel onderscheiden in toetsingsvorm, bestuderen de leerlingen bij de diagnostische toets (experiment I) de tekst duidelijk korter dan de leerlingen bij de summatieve toets (experiment II), en hebben ze bovendien een aanmerkelijk lagere antwoordscore, met name bij de eerste, pittige vwo-set. Dit sluit aan bij de *parsing the context hypothesis* (Britt et al., 2018): naarmate het belang van de taak minder groot is, zal een lezer zich minder snel en minder diepgaand tot een complexe leestaak zetten. Een diagnostisch gebruikte toets zet een lezer dan – het ligt voor de hand – minder aan tot een diepe, nauwkeurige, gelaagde werkwijze dan een toets met een summatief doel.

4.1 Kanttekeningen

Natuurlijk zijn er bij deze studie de nodige voorbehouden te plaatsen. Vooropgesteld: dit onderzoek had uitsluitend betrekking op (tamelijk pittige) situatiemodelvragen, bij relatief lange, zakelijke opiniërende teksten, binnen de taakcontext van een (al dan niet diagnostische) leestoets, onder vaardiger (adolescent-)lezers. Andersoortige vragen (eenvoudigere of globalere vragen, analytische vragen, meer tekstbasis-gebonden vragen, etc.), bij andere teksten (kortere zakelijke teksten, andere genres zakelijke teksten, fictionele teksten, etc.), in een andere taakcontext (formatieve toetsing, 'gewone' leestaak in de les, etc.), onder andere lezers vmbo-leerlingen, volwassen expert-lezers, etc.) zouden mogelijk andere resultaten hebben opgeleverd.

Daarnaast onderzochten we in deze studie de intensiteit van tekstbestudering, niet zozeer de kwaliteit ervan: het valt niet uit te sluiten dat een onderscheid binnen de steekproef tussen vaardiger en minder vaardige lezers andere accenten in de onderzoeksresultaten zou hebben opgeleverd. Verder dient opgemerkt dat leerlingen in onze studie de vragen vooraf niet konden inzien: wanneer ze deze kort vooraf hadden kunnen inzien voordat ze de tekst bestudeerden, hadden we mogelijk andere resultaten gevonden. Wel is het de vraag of vaardigere lezers bij een taak als deze vanuit zichzelf eerst de vragen kort doornemen: in conditie IIC 'Vragen' hadden de vwo 5-leerlingen hiertoe de mogelijkheid maar startten ze desondanks doorgaans met het doornemen van de tekst.

Bovenal: de complexiteit, lokaliteit en aard van de vraagstelling én de complexiteit van de teksten zelf zijn waarschijnlijk zeer bepalend voor de resultaten in deze

studie. Doordat we bij twee tamelijk complexe teksten situatiemodelvragen over de centrale gedachtegang stelden, werden hoge eisen aan het tekstbegrip van leerlingen gesteld. Daardoor zullen leerlingen vaak de behoefte hebben gehad de tekst nogmaals te raadplegen. Eenvoudigere, globalere vragen bij eenvoudiger teksten zullen lezers waarschijnlijk vaker vanuit hun vooraf-lezing van de tekst kunnen beantwoorden.

4.2 Implicaties voor het onderwijs

Deze studie heeft interessante implicaties voor het onderwijs in tekstbegrip. In Nederland staat de tekst met vragen als leestaak al sinds lange tijd ter discussie, omdat ze van een leerling een ‘weinig kritische lezer’ zou maken die ‘niet zelfstandig complexe teksten kan aanpakken zonder dat er vragen worden gesteld’ (Van de Wetering & Groenendijk, 2015, p. 5). Al bestaat hierover een brede consensus onder onderwijskundigen, docenten, taalwetenschappers en leesonderzoekers (bijv. Hoogeveen, 2018; LTN, 2019b; SLO, 2017), empirische onderbouwing bij deze kritiek ontbrak. Onderhavige studie impliceert dat een diepgaande, zelfstandige tekstbestudering vooraf voor leerlingen bij deze taak hoogstwaarschijnlijk geen meerwaarde heeft. Daarmee wordt de kritiek aannemelijk dat leerlingen niet zozeer met deze taak leren zelfstandig een tekst te bestuderen maar dat ze vooral ermee leren vragen over een tekst correct te beantwoorden.

Binnen de context van de gemiddelde leesles Nederlands wordt dat laatste zelfs nog aannemelijker. In onze studie wordt weliswaar materiaal uit de CSE's Nederlands havo/vwo gebruikt, we handhaafden daarin alleen de situatiemodelvragen over de centrale tekstuele gedachtegang. In veel huidige leergangen Nederlands voor het Nederlandse secundaire onderwijs worden in leestaken echter sterk uiteenlopende vraagtypen gebruikt, in een nogal onvoorspelbare afwisseling: situatiemodelvragen staan tussen functioneel-analytische vragen, vragen over vocabulaire en argumentatie-analytische en -beoordelende vragen (Roosjacks et al., 2021a). Een dergelijke taakinrichting bevordert nog sterker taakafhankelijk leesgedrag. Leerlingen hebben dan immers vooraf geen idee wat hun gevraagd zal gaan worden.

Dat leerlingen via deze reguliere invulling van de tekst met vragen vooral vragen over een tekst leren te beantwoorden, spreekt vanuit wetenschappelijk oogpunt overigens zo goed als vanzelf. Zonder duidelijke taakcontext, zonder concreet doel wordt de vooraf-tekstbestudering vooral een middel om het concrete, meetbare doel in de taak te bereiken: het correct beantwoorden van de vragen. Dat heeft twee logische gevolgen voor het lezersgedrag bij dit soort taken. Ten eerste, omdat vooraf-lezen vooral een hulpmiddel is om de vragen beter te begrijpen en te beantwoorden, lezen leerlingen vooraf nauwelijks met diep begrip (vgl. Onderwijsraad & Raad voor Cultuur, 2019; zie ook Wolf, 2018) en lezen ze vooral ‘oppervlakkig’: leerlingen bouwen met name een tekstbasis op en enkel een zeer globaal situatiemodel. Dergelijk oppervlakkig leesgedrag werd inderdaad bij vwo-leerlingen aangetroffen (hoofdstuk 2 en 3; Breukink, In voorbereiding; Stevens, 2018). Ten tweede, doordat

leerlingen zich vooral richten op de vragen, zullen ze vaak een hoge mate van taakvertrouwdheid en -afhankelijkheid opbouwen. Leerlingen worden zowel in primair als in secundair onderwijs vaak aan deze leestaak blootgesteld, en zo bouwen ze gedetailleerde kennis op over vraagtypen, en eveneens over vaardigheden en strategieën die per vraagtype wenselijk én adequaat zijn. Sporen van deze taakvertrouwdheid vonden we in onze eerdere studie in hoofdstuk 3.

Betekent dat nu dat een tekst met vragen geen tekstbegrip toetst? Voor de goede orde: wanneer leerlingen een vraag over een tekst goed beantwoorden, geven ze natuurlijk wél blijk van een vorm van tekstbegrip. In het leven van alledag hebben leerlingen immers vaak te maken met schoolse situaties waarin ze met een gerichte zoekvraag een tekst tot zich nemen, bijvoorbeeld bij werkstukken – zoals er ook tal van schoolse situaties zijn waarin ze vragen over een tekst moeten beantwoorden zónder dat de brontekst nog raadpleegbaar is, bijvoorbeeld bij proefwerken voor zaakvakken (Artelt et al., 2001; Ferrer et al., 2017; Johnston, 1984; Schroeder, 2011). Deze beide vormen van tekstbegrip (vragen beantwoorden met én zonder de tekst erbij) lijken redelijk te correleren (Schroeder, 2011). Daarom kunnen de resultaten in deze studie in onze ogen ook niet worden gezien als ondersteuning bij de stelling dat een tekst met vragen slechts in beperkte mate begrip op het situatiemodel toetst (vgl. Kamalski, 2007), al sluiten de resultaten wel aan bij recente bedenkingen over de beperkte cognitieve validiteit van de tekst met vragen als leestoets: de begripsprocessen bij het beantwoorden van vragen bij een tekst overlappen maar ten dele met de begripsprocessen van leerlingen bij ‘echte’ leestaken (zie met name O’Reilly et al., 2018). Implicaties van deze studie liggen, denken we, veeleer op inhoudelijk-didactische inrichting van de leesles Nederlands, met name ten aanzien van de tekst met vragen als trainingsinstrument. We werken er vier uit.

De eerste implicatie: vermijd een taakinrichting met vragen die inhoudelijk sterk uiteenlopen en elkaar onvoorspelbaar opvolgen en die niet aan een duidelijk doel of context zijn gekoppeld. Een andere werkwijze is aan te bevelen. Door bij deze taak te werken met een duidelijk, concreet leesdoel vooraf en met vragen die logisch aansluiten op dit concrete leesdoel, wordt de vooraf-bestudering van de tekst voor de lezer in een duidelijke leessituatie ingebed (McCrudden et al., 2010; O’Reilly et al., 2018). Zo’n leesdoel kan bijvoorbeeld een voorbereidende leestaak voor een schrijfoopdracht zijn, maar kan ook worden aangeboden binnen een fictief, voor leerlingen herkenbaar scenario, van waaruit ze vooraf de tekst moeten lezen en daarna begripsvragen moeten beantwoorden (OECD, 2019; Sabatini et al., 2018). Eventueel kan dan, zoals in onze studie, de tekst worden weggehaald bij het overhandigen van de vervolgvragen, waardoor de vragen voor leerlingen een check achteraf worden of ze de tekst vooraf voldoende hebben begrepen (Schroeder, 2011).

Een tweede implicatie: spreid in taaksoorten. Door in de leesles een veelvoud aan soorten taken aan te bieden (sorteertaken, argumentatietaken, synthesetaken, etc.), kunnen leerlingen minder taakafhankelijk en taakvertrouwd gedrag opbouwen. Als ze daarvoor de tekst vooraf zelfstandig moeten bestuderen, dan moet voor leerlingen wel vooraf duidelijk zijn aan welke dimensie van de tekst ze bij het lezen

van de tekst speciale aandacht moet schenken: bijv. centrale gedachtegang, aanvaardbaarheid argumentatie, inhoudelijke consistentie, etc.

Een derde implicatie: stel bij leestaken vaker het leesproces zelf centraal. De aandacht gaat in de les zakelijk lezen nu vaak uit naar het leesproduct, doordat vragen bij een tekst al dan niet klassikaal worden beantwoord en besproken. Deze studie impliceert dat leerlingen voor een goede prestatie bij een tekst met vragen zich het beste kunnen richten op het antwoordproces, en niet zozeer op het vooraf-leesproces; daardoor raakt het zelfstandig bestuderen van een complexe tekst al snel buiten beeld. En weliswaar *modellen* docenten vaak, doen docenten voor hoe ze zelf zich teksten eigen maken (bijv. Linthorst & De Glopper, 2015), ze lijken in de les leerlingen zelden ertoe aan te zetten op hun eigen leesgedrag te reflecteren. In de leesles zou daarom bijvoorbeeld vaker het leesproces van leerlingen zelf centraal kunnen staan: hoe doorgronden leerlingen zelfstandig complexe teksten, welke leesproblemen ervaren ze dan en vanuit welke vaardigheden en strategieën kunnen ze nog beter dit soort teksten begrijpen (vgl. Couzijn, 1995; Rooijackers et al., 2021b)?

Een vierde, meer indirecte implicatie is: verruim de blik op tekstbegrip in de les zakelijk lezen. In deze studie ontwierpen we twee experimenten waarbij leerlingen één samenhangende, kwalitatief hoogwaardige tekst via vragen gaandeweg moesten doorgronden, en daarmee sloten we aan bij de courante taakinrichting in huidige leidende leergangen Nederlands (Rooijackers et al., 2021a). Vakdidactische rationale achter deze taakinrichting is de gedachte dat leerlingen via vragen de diepere gedachtegang in een complexe betrouwbare tekst blootleggen en door oefening gaandeweg zichzelf deze benadering eigen maken (Hendrix & Hulshof, 1994). Daarmee blijven echter theoretische inzichten in tekstbegrip buiten beschouwing van met name ná de millenniumwisseling, inzichten die in het theoretisch kader van onze studie zijn meegenomen. Met name relevant lijkt het inzicht dat veel tekstbegrip in het leven van alledag wordt opgebouwd vanuit een specifiek leesdoel (Britt et al., 2018; vgl. McCrudden et al., 2011). Door deze nadruk op de context rond de tekst en de lezer, komen andere vaardigheden bij tekstbegrip in beeld: het kunnen bepalen van de kwaliteit, betrouwbaarheid en bruikbaarheid van een tekst, en het kunnen opbouwen van adequate taak- en contextmodellen bij leestaken (Britt & Rouet, 2012; Gubbels, 2020). Deze aspecten worden nu vaak niet of nauwelijks in de gemiddelde leestaak van het schoolvak Nederlands meegenomen (Gubbels, 2020; Rooijackers et al., 2021a), wat een van de verklaringen kan zijn waarom in internationaal onderzoek Nederlandse 15-jarige leerlingen relatief laag scoren op het domein reflecteren en evalueren (Gubbels et al., 2019).

Praktisch zorgt zo'n blikverruiming voor andere aandachtsgebieden in de leesles Nederlands (zie Rooijackers et al., 2021a). Leerlingen oefenen met het opbouwen en toepassen van adequate context- en taakmodellen in complexe leessituaties en reflecteren hierop: denk aan taken zoals werkstukken waarbij leerlingen informatie uit meerdere bronnen synthetiseren (Gil et al., 2010). Of leerlingen werken met leestaken waarin de teksten sterk variëren in kwaliteit, betrouwbaarheid en bruikbaarheid; daarbij ontwikkelen ze dan criteria en passen deze vervolgens toe op de

set teksten (Anmarkrud et al., 2014). Of de aandacht gaat in de les uit naar de selectiecriteria van informatiebronnen op internet (Bråten et al., 2017).

4.3 Conclusie

In deze studie onderzochten we via twee experimenten de relatie tussen het vooraf bestuderen van een tekst en de antwoordscore, bij twee teksten met vragen. Blijkens het eerste experiment komen de uiteindelijke antwoorden grotendeels tot stand doordat deelnemers tijdens de vragen de tekst kunnen raadplegen. De niet-betekenisvolle uitkomsten uit het tweede experiment lijken erop te wijzen dat leerlingen door vooraf intensief de tekst te bestuderen nauwelijks tot niet beter presteren bij deze taak.

Echter, er is nog de nodige ruimte voor vervolgonderzoek. Vervolgonderzoek zou bijvoorbeeld kunnen verkennen hoezeer de uitkomsten van dit onderzoek afhangen van de steekproef (vaardige lezers), het soort teksten (complexe opiniërende zakelijke teksten) en het vraagtype (situatiemodelvragen). Daarnaast zou vervolgonderzoek ook de relatie tussen de vooraf-tekstbestudering en de antwoordscore bij andere taken (sorteertaken, etc.) kunnen onderzoeken.

HOOFDSTUK 5

Observerend leren lezen op het vwo

*Reflecteren op tekstbegrip via oogbewegingsfilmpjes**

* Dit hoofdstuk is in een eerdere vorm verschenen als: Rooijackers, P., Silfhout, G. van, & Bergh, H. van den. (2021). Observerend leren lezen op het vwo; reflecteren op tekstbegrip via oogbewegingsfilmpjes. *Levende Talen Tijdschrift*, 22(3), 3-16.

Samenvatting

Observerend leren heeft vrij breed ingang gevonden in het schrijfonderwijs in het schoolvak Nederlands. In het leesonderwijs gebeurt het via *modellen* op expert-niveau: een docent doet leerlingen voor hoe hij een tekst al lezend begrijpt. *Modellen* op *peer*-niveau, via het leesgedrag van leerlingen, lijkt nog sporadisch te gebeuren. In onze studie hebben we bij vwo 3- en vwo 4-leerlingen ($N = 131$) onderzocht in hoeverre het proces achter diep tekstbegrip effectief kan worden onderwezen via oogbewegingsfilmpjes (*EMME*'s) van lezende leerlingen, teneinde het tekstbegrip te bevorderen. De deelnemers aan de *EMME*-lessenserie geven achteraf meer van diep begrip blijk dan de leerlingen die zelf teksten lezen en vragen maakten.

1 Inleiding

1.1 Aanleiding

Er is nogal wat onderzoek dat de effectiviteit van observerend leren in de les Nederlands laat zien (o.a. Braaksma, 2002; Couzijn, 1995). Een leerling hoeft niet direct zelf een taak uit te voeren maar observeert eerst de taakuitvoering of het product van een model, reflecteert erop en evalueert het. Omdat hij de taak niet zelf uitvoert, heeft hij meer cognitieve ruimte beschikbaar om te leren: hij vergaart zo meta-cognitieve en procedurele kennis over de taakuitvoering en kan deze vervolgens toepassen als hij zelf een soortgelijke taak gaat verrichten (Bandura, 1986). Vooral binnen het domein schrijfvaardigheid is deze positieve werking van observerend leren tamelijk uitgebreid in kaart gebracht: leerlingen blijken in een groot aantal studies betere teksten te produceren wanneer ze vooraf hebben nagedacht over de uitvoering van eigen of andermans schrijfproduct, dan wanneer ze vooraf zelf een schrijftaak hebben verricht (Bouwer et al., 2018b; Braaksma, 2002; Couzijn, 1995; Elving-Heida, 2019; Rijlaarsdam et al., 2008; etc.).

Aanmerkelijk minder vaak bestudeerd is de effectiviteit van observerend leren voor de leesles Nederlands (Couzijn, 1995; Keehnen et al., 2015). Deze beperktere wetenschappelijke aandacht kan met meerdere factoren samenhangen. Lezen zou minder cognitief belastend zijn dan schrijven (Kellogg, 1994). Daardoor hebben leerlingen bij lezen mogelijk minder behoefte aan het observeren van voorbeeldgedrag dan bij schrijven. Bovendien is het niet altijd eenvoudig voor een toeschouwer goed inzicht te krijgen in het leesproces van lezers. (Jonge) lezers reflecteren met wisselende scherpthe op hun eigen leesgedrag en zijn zich vaak niet bewust van kleinschalige begripsreparaties, zoals bij een moeilijk woord of een ingewikkelde zinsnede (Holmqvist & Andersson, 2017).

Observerend leren heeft op dit moment ook slechts beperkt ingang gevonden in de gemiddelde leesles Nederlands: leerlingen gaan vooral zelfstandig aan de gang met vragen bij teksten, waarna ze, eventueel in tweetallen, hun antwoorden nakijken en bespreken, of ze luisteren naar een docentbespreking hiervan (Linthorst & De Gloppe, 2015). Het leesgedrag van leeftijdsgenoten lijkt maar zelden in de gemiddelde leesles centraal te staan; vooral de observatie van expert-leesgedrag van docenten, *modellen*, lijkt een duidelijke plaats te hebben veroverd (Bogaerds-Hazenbergh & Evers-Vermeul, 2020).

Het lijkt echter zinvol observerend leren ook vanuit *peer*-niveau toe te passen in de leesles Nederlands. Vaardige leerlingen blijken bij schrijfvaardigheid vooral te leren door expertvoorbeelden te observeren terwijl minder vaardige lezers vooral leren door *peer*-voorbeelden te bekijken, die voor hen meer herkenbare problemen met de taak laten zien (Braaksma, 2002; vgl. Schunk, 1987). In deze studie hebben we de effectiviteit van observerend leren lezen op *peer*-niveau onderzocht, niet via video's zoals eerder Couzijn (1995) en Keehnen et al. (2015) deden, maar via oogbewegingsfilmpjes (*EMME's: Eye-Movement Modeling Examples*).

1.2 Theorie en onderzoeksvragen

Allereerst: wat is tekstbegrip? Volgens de Constructie-Integratietheorie van Kintsch (1998) is een tekst een netwerk van hiërarchisch samenhangende, betekenisvolle beweringen. Steeds samenhang in dit netwerk aanbrendend, leggen lezers verbindingen tussen woorden en zinnen (een microstructuur) en verbindingen tussen alinea's (een macrostructuur). Zo bouwen ze een 'tekstbasis' op, een vrij letterlijke mentale representatie van de tekst, en een 'situatiemodel', een mentaal model waarin ze de tekstuele wereld integreren in hun eigen voorkennis, ideeën etc. Pas op dit niveau kennen ze aan een tekst diepere betekenis toe.

Om de totstandkoming van tekstbegrip te kunnen beschrijven, hebben leesonderzoekers de laatste drie decennia steeds vaker gebruikgemaakt van eye-tracking. Eye-tracking registreert waarop en hoelang lezers hun ogen fixeren. Verschillen in de duur van oogfixaties zijn indicatief voor verschillen in cognitieve verwerkingstijd. Eye-tracking biedt een nauwkeurige weerspiegeling waarnaar lezers in een tekst kijken; daarmee laat het vooral zien waaraan lezers in een tekst aandacht schenken, maar niet waarom zij daaraan aandacht schenken (Holmqvist & Andersson, 2017). Het laatste decennium worden daarom wel eye-tracking en hardop-denken in leesonderzoek gecombineerd, zodat zowel het 'wat' als het 'waarom' achter leesgedrag kan worden beschreven (hoofdstuk 2 en 3; Bax, 2013).

Eye-tracking is in deze *mixed method*-vorm ook inzetbaar in de leesles, via EMME's. Een proefpersoon voert een taak uit op een beeldscherm, terwijl een eye-tracker zijn oogbewegingen (fixaties) op het beeldscherm registreert. Een onderzoeker construeert later van deze oogbewegingen op het scherm een filmpje en monteert het audiocommentaar van de proefpersoon eronder. Deze EMME kunnen docenten vervolgens in de les gebruiken om op leeractiviteiten te reflecteren. Afbeelding 5.1 geeft een beeld wat een leerling bij een EMME te zien en te horen krijgt. (Een voorbeeld van een EMME uit onze studie is te bekijken op YouTube: <https://youtu.be/mkW1UqQjKmk>.) EMME's bieden duidelijke didactische voordelen: leerlingen krijgen een vrij directe inblik in de complexe perceptuele en cognitieve processen van een lezer (Van Marlen, 2019). Bovendien blijkt het bekijken van EMME's vaak motiverend (Salmerón & Llorens, 2019). Veel internationale didactische studies vonden dan ook positieve leereffecten bij EMME-inzet in de klas (zie voor een overzicht: Van Marlen, 2019).

BoukjeAI2

Complotdenkers

[1] Op onze aarde lopen een hoop complotdenkers rond die de meest bizarre dingen beweren. Zo was de maanlanding in 1969 in scène gezet en zijn er plannen om Europa de komende decenia helemaal Duits te maken. Die theorieën zijn natuurlijk best leuk maar sommige gaan wel erg ver. Onderzoekers braken zich lange tijd het hoofd over wat sommige mensen wel complotdenkers maakt en anderen niet. Nu denkt een team psychologen van de Vrije Universiteit het antwoord te weten: complotdenkers gebruiken een bepaald denkpatroon ('illusoire patroonherkenning') overduidelijk veel.

[2] Het team kwam tot deze conclusie na het uitvoeren van verschillende experimenten op 264 volwassenen. Zo testten de psychologen eerst of veelrumpers verzochten complottheorieën geklemd op een schaal van 1 tot 5. Vervolgens onderzochten de wetenschappers of deelnemers sneller patronen zagen in willekeurige dingen, zoals een muntje opgooien. Mensen die bij het eerste onderzoek hoog scoorden, waren ook geneigd hierin patronen te zien. Deze bevindingen werden bevestigd in nog drie experimenten.

[3] De conclusie die het team uit deze experimenten trok, is dat complotdenkers te veel en te vaak verbanden zien tussen gebeurtenissen die er eigenlijk niet zijn: 'illusoire patroonherkenning'. Iedereen heeft dat wel eens. Je droomt bijvoorbeeld over je opa en de dag erna hoor je dat hij is overleden. Dit kan haast geen toeval zijn, denk je dan. Maar complotdenkers gaan in dit ontdekken van patronen te ver.

[4] Sociaal psycholoog Michiel van Elk merkt naar aanleiding van dit onderzoek op: 'Psychologen onderzoeken al langer de denkwijze van complotdenkers. Veel wetenschappers hebben geen verband gevonden, in dit onderzoek hebben ze dat wel. De vraag is hoe sterk de aangetoonde relatie is. Toevalstreffers kun je immers niet uitsluiten. En andere factoren blijven volledig buiten beschouwing.'

0:22 / 0:39 Scroll voor details

Afbeelding 5.1 Een fragment uit een Eye-Movement Modeling Example

Een kijker ziet over de woorden op het scherm een klein bewegend bolletje lopen dat de oogfixaties van de proefpersoon representeert en hoort er haar commentaar bij. De grootte van het bolletje geeft de lengte van de fixatie op een woord weer.

In dit voorbeeld blijft de lezer hangen bij elke zinnen in alinea 2 en levert hierbij als commentaar: "Ja, ik vond dat hier veel informatie tegelijk gegeven werd. En nou ja ... ik probeerde goed te begrijpen hoe ze nou precies dat onderzoek hadden uitgevoerd. Dus daarom lees ik hier de dingen opnieuw. Maar het kwam niet binnen ..." Interviewer vraagt: "Zet de alinea te snelle stappen?" "Ja ... maar door terug te lezen snapte ik de alinea nu wel beter."

Binnen het domein van digitale leesvaardigheid is de werking van EMME's inmiddels wetenschappelijk verkend. In de studie van Salmerón en Llorens (2019) observeerden derde klas-leerlingen uit het voortgezet onderwijs via EMME's leeftijdsgenoten die strategieën toepasten bij digitale leestaken, zoals het plannen van de aanpak, het selecteren en evalueren van relevante informatie en het monitoren van de aanpak. En Salmerón et al. (2020) lieten leerlingen EMME's zien van leeftijdsgenoten die in wisselende mate van effectiviteit de betrouwbaarheid van informatie op internet beoordeelden. In beide studies pasten de leerlingen in de experimentele conditie vervolgens vaker de eerder geobserveerde strategieën toe dan leerlingen in de controleconditie. Salmerón en Llorens (2019) vonden ook een verhoogd tekstbegrip op de natoets bij deze leerlingen, Salmerón et al. (2020) echter niet.

Nederlands onderzoek naar observerend leren lezen maakte tot nu toe gebruik van video's waarin leerlingen hardop denkend een tekst proberen te begrijpen.

In Couzijn (1995) zagen havo 4-/vwo 4-leerlingen leeftijdsgenoten in een video hardop argumentatieve leestaken verrichtten en hoorden ze hoe deze in wisselende mate van effectiviteit leesstrategieën toepasten, zoals zich op de taak oriënteren en het proces monitoren. En in de verkennende studie van Keehnen et al. (2015) bekeken vwo 3-leerlingen gedurende één les een video van een vwo 6-leerling die hardop een tekst met passende leesstrategieën doorgrondde en vragen erbij maakte. In beide studies scoorden leerlingen in de experimentele conditie hoger op een tekstbegripstoets, dan leerlingen in de controleconditie.

Nog meer dan video's echter, lijken EMME's een geschikt middel om in de leesles leerlingen op het opbouwen van tekstbegrip te laten reflecteren. Leerlingen kunnen bij EMME's lezers immers niet alleen horen praten over hun leesgedrag, maar kunnen ook zien aan welke tekstdelen deze aandacht schenken. Beeld en commentaar vullen elkaar aan, waardoor leerlingen makkelijker het voorbeeldgedrag kunnen volgen en interpreteren; bovendien bepaalt het reflectievermogen (*self-explaining*) van de modellen aanzienlijk minder in hoeverre hun gedrag is te volgen (Van Marlen, 2019) *EMME's* staan zo een breder palet aan navolgbare modellen toe dan video's.

De laatste jaren weerklinkt herhaaldelijk de kritiek dat Nederlandse leerlingen onvoldoende leren diep te lezen, oftewel 'geconcentreerd te lezen, waarbij de samenhang en betekenis van een tekst worden ervaren' (Onderwijsraad & Raad voor Cultuur, 2019). Wolf (2018) wijst internationaal op de noodzaak om 21e-eeuwse scholieren in de gedigitaliseerde samenleving nadrukkelijker te scholen in diep begrip. Er is in Nederland ook een duidelijke roep om meer *close reading* in de leesles waarneembaar (bijv. Lapp et al., 2018). Deze roep is waarschijnlijk niet zonder reden: het tekstbegrip van Nederlandse 15-jarige leerlingen vertoont een dalende trend, waarbij leerlingen vooral minder scoren op het subdomein reflecteren en evalueren (Gubbels et al., 2019). Tekenend in dit opzicht lijkt het leesgedrag van vaardigere lezers in het voortgezet onderwijs, vwo-leerlingen: wanneer zij de expliciete instructie tot een intensieve tekstbestudering krijgen, blijken ze vaak te volstaan met een nogal oppervlakkig, sterk aan de tekstbasis gebonden begrip (zie hoofdstuk 2 en 3; vgl. Breukink, In voorbereiding).

Er lijkt didactisch behoefte aan meer instrumenten om met leerlingen op (de kenmerken van) diep begrip te kunnen reflecteren. Observerend leren lezen lijkt hiertoe bij uitstek geschikt. In deze studie hebben we daarom de effectiviteit van observerend leren voor diep begrip nader bekeken via EMME's, waarbij we ons richten op vwo 3- en vwo 4-leerlingen. Daarbij maken we gebruik van (uiteenzetende / betogende) zakelijke teksten uit kranten en tijdschriften. We vragen ons in deze studie af: a) In hoeverre kan het observeren van EMME's vwo-leerlingen bewustmaken van de voorwaarden voor diep begrip bij zakelijke teksten? En b) in hoeverre geven leerlingen die eerder de EMME's hebben geobserveerd, blijk van meer diep begrip als ze daarna zelf zakelijke teksten lezen?

2 Methode

2.1 Experimentopzet en participanten

Om de effectiviteit van de EMME-interventie te testen, hebben we gebruikgemaakt van een *pretest-posttest control-group design*. Het (quasi-)experiment is uitgevoerd in drie vwo 3- en drie vwo 4-klassen op één Nederlandse middelbare school ($N = 131$). Daarbij is per leerjaar telkens één klas aan de controleconditie en zijn de twee andere klassen aan de experimentele conditie toegekend.

De experimentele lessenserie bestond uit vijf lessen van 50 minuten: twee EMME-lessen, afgewisseld met drie werklessen met speciaal ontworpen leestaken. In de controleconditie ontvingen de leerlingen onderwijs in tekstbegrip vanuit de bekende leergang *Nieuw Nederlands*: in twee lessen werden klassikaal de in deze leergang opgenomen 'leesstrategieën'¹ besproken en toegepast, vooral via *modeling* op expert-niveau; in drie werklessen maakten leerlingen teksten met vragen. Alle lessen zijn gegeven door de betrokken docenten Nederlands van de school, daarin begeleid door een van de onderzoekers.

2.2 Doelen experimentele lessenserie

We hanteerden bij de experimentele lessenserie twee ontwerpprincipes (doelen). Ten eerste wilden we in de hele lessenserie aansluiten bij een conceptualisatie van diep begrip à la Kintsch (1998): pas als een lezer een adequaat situatiemodel opbouwt, is er diep begrip. Leerlingen moesten zich ervan bewust worden dat diep begrip zich niet zozeer kenmerkt door een oppervlakkige, woordelijke tekstherinnering, als wel door een actieve verwerking van de tekst, in eigen woorden, tijdens én na lezing. Dat namen we niet alleen in de twee EMME-lessen als uitgangspunt, maar ook in de drie werklessen. Daarin werden alle leestaken zo geconstrueerd dat leerlingen vooraf de tekst intensief moesten bestuderen. Vooraf vernamen ze namelijk dat ze bij de vervolgtask niet langer over de tekst zouden gaan beschikken. Wanneer lezers weten dat ze bij de vervolgtask niet langer de beschikking hebben over de tekst, blijken ze deze langer en met meer oog voor diepere betekenis te bestuderen (Schroeder, 2011).

Ten tweede wilden we in de twee EMME-lessen leerlingen laten inzien dat ze vaak tekstdelen moeten herlezen als ze een complexe tekst willen doorgronden. We wilden leerlingen geen vaste, eenduidig inzetbare set leesstrategieën aanbieden: hoe leesstrategieën worden ingezet tijdens het lezen, is immers persoons-, tekst- en doelafhankelijk (Almasi & Fullerton, 2012; Rooijackers et al., 2021b). Wanneer twee lezers eenzelfde tekst 'strategisch' lezen, plaatsen ze andere accenten in een tekst en passen daarbij (deels) andere strategieën toe. Intensief teruglezen is echter een overkoepelend kenmerk van veel leesstrategieën: door zinnen, regels

¹ In *Nieuw Nederlands*, 6e editie, worden leesstrategieën onzes inziens weinig evidencebased uitgewerkt en aangeboden. Zie Rooijackers et al. (2021b).

en alinea's terug te lezen, kan een lezer zowel falend tekstbegrip repareren als zijn begrip van een tekst controleren en verdiepen (Van den Broek & Helder, 2017).

2.3 Constructie materiaal experimentele lessenserie

Constructie EMME-lessen. Vanuit het eye-track-experiment waarover we rapporteerden in hoofdstuk 3, beschikten we over een ruime hoeveelheid eye-track-data van 56 vwo-leerlingen die korte uiteenzettende of betogende teksten (achtergrondartikelen of opiniestukken) lazen uit Nederlandse dagbladen en tijdschriften. Daarnaast beschikten we over retrospectieve interviews waarin leerlingen de eye-trackbeelden van hun lees- en antwoordgedrag bekeken en hierop reflecteerden. De data bij twee teksten uit dit experiment dienden als basis voor de constructie van de twee EMME-lessen.

In de EMME-literatuur wordt benadrukt dat EMME's aan leerlingen als een onderling sterk contrasterende set moeten worden aangeboden (Salmerón & Llorens, 2019). Daarom selecteerden we bij deze twee teksten twee drietallen van eye-track-filmpjes (*scan-paths*). Binnen deze drietallen onderscheidde vaardige lezers en minder vaardige lezers zich duidelijk, in leestempo (want vaardige lezers zijn doorgaans snellere lezers dan minder vaardige lezers: Perfetti & Stafura, 2014) en in strategiegebruik (vaardige lezers zetten doorgaans vaker en efficiënter strategieën in die het opgebouwde begrip bij een tekst verdiepen: Van den Broek & Helder, 2017). Om dit contrast tussen vaardige en minder vaardige lezers nog te bevorderen, besloten we niet gehele filmpjes te gebruiken, maar selecteerden we fragmenten van doorgaans een anderhalve minuut, waarin lezers zich duidelijk in leesgedrag van elkaar onderscheidde. Zo namen we in het eerste trio een vaardige leerling op die, nadat hij eerst de gehele tekst snel en efficiënt had doorgewerkt, alle kernzinnen trefzeker lokaliseerde en herlas, waarmee hij zijn begrip van de tekst duidelijk verdiepte. Daartegenover stond een minder vaardige leerling die nogal langzaam las en ondertussen slechts enkele woordgroepen herlas, waarmee hij vrijwel uitsluitend lokale begripsproblemen repareerde. De derde, 'gemiddeld vaardige' lezer las relatief langzaam, herlas veel zinnen, met veel kleine en middelgrote reparaties, en ze herlas op het laatst nog enkele kernzinnen: deze lezer repareerde en verdiepte haar begrip. Een screeningspaneel van drie docenten Nederlands beoordeelde de twee trio's vooraf op bruikbaarheid.

Vanuit het oorspronkelijke audiocommentaar ontwikkelden we vervolgens scripts om de beelden te begeleiden. In het oorspronkelijke commentaar blikten leerlingen terug op eye-trackbeelden van hun leesgedrag: begrepen ze wat ze lazen, waarom keken ze terug naar eerdere zinnen of tekstdelen? Dit commentaar bleek echter technisch van een te lage kwaliteit en inhoudelijk moeilijk te volgen voor buitenstaanders. Daarom construeerden we op basis van de oorspronkelijke leerlinguitspraken scripts. Daarin manipuleerden we een tweetal zaken. Ten eerste de mate waarin modellen de tekstinhoud in eigen woorden adequaat weergaven: een vaardige lezer herhaalt niet letterlijk tekstbewoordingen, maar geeft de tekstinhoud in eigen woorden weer. En ten tweede de kwaliteit waarmee modellen re-

flecteerden: een vaardige lezer kan scherper het rationale achter eigen leesgedrag verwoorden dan een minder vaardige lezer (McNamara & Magliano, 2009b). Na screening spraken de vwo 4-leerlingen deze scripts vervolgens in. Ten slotte werden de videobeelden en het audiocommentaar gesynchroniseerd tot een EMME.

In een aantal try-outs van de EMME-lessen, in respectievelijk één vwo 2-, één vwo 3 en één vwo 4-klas, werd een drietal zaken gecontroleerd: of het materiaal niet te cognitief belastend was voor leerlingen, of de gescripte modellen voldoende realistisch en herkenbaar waren voor de doelgroep en in hoeverre het materiaal ongewenste reacties van leerlingen uitlokte. Het eerste en tweede punt bleken geen problemen op te leveren; het derde punt deels wel. Twee van de zes EMME's bleken te multi-interpretabel en werden daarom vervangen.

Bij het samenstellen van de leerlingtaken voor de twee EMME-lessen werd sterk erop gelet dat aan de voorwaarden van observerend leren werd voldaan. Leerlingen moesten tijdens de twee lessen niet alleen op het leesgedrag van de aangeboden modellen reflecteren (Bandura, 1986), maar dit ook evalueren (Braaksma, 2002). Leerlingen analyseerden daarom niet alleen de EMME's afzonderlijk, maar ze moesten ook beargumenteren welk model, op basis van de vertoonde beelden en het hoorbare commentaar, naar hun idee het duidelijkst diep begrip opbouwde, en welk model het minst. In een klassengesprek inventariseerde en evalueerde de docent ten slotte de individuele bevindingen van leerlingen.

Constructie materiaal experimentele lessenserie: werklessen. Voor de drie werklessen, volgend op elke EMME-les, werden vier korte uiteenzettende of betogende teksten uit Nederlandse dagbladen of tijdschriften geselecteerd. Hierbij werden vier 'situatiemodeltaken' ontwikkeld: (1) twee taken met losse vragen op situatiemodelniveau over de tekst, (2) een sorteertaak en (3) een samenvattingstaak. We voegden de instructie toe dat de tekst ná bestudering, bij het uitdelen van de vervolgtask, zou worden ingenomen. We wilden namelijk dat leerlingen al tijdens de tekstbestudering diep begrip opbouwden, en niet pas tijdens het verrichten van de vervolgtask. Leerlingen beschikten dus bij de vervolgtask niet langer over de teksten.

2.4 Voor- en nameting

Om te kunnen controleren op verschillen in tekstbegrip tussen de experimentele en controleconditie, hanteerden we een pretest-posttest-design. Als voortoets werd een cloze-toets gebruikt: een gatentekst, waarin woorden strategisch zijn weggelaten (vgl. Oller & Jonz, 1994). In de posttest toetsten we diep begrip via twee leestoetsen: een cloze-taak en een reguliere tekst met vragen, met vooral open vragen en enkele gesloten vragen. Deze beide toetsvormen kunnen uiteenlopende aspecten van diep tekstbegrip bevragen (Kamalski, 2007).

Omdat afname onder vwo 3- of vwo 4-leerlingen vanwege de Covid19-epidemie niet mogelijk bleek, werden de twee natoetsen gepretest onder vwo 5-leerlingen op één middelbare school ($N = 139$). Op grond van de resultaten werden items met lage item-rest-correlaties uit de toetsen verwijderd, evenals items met een te hoge

moeilijkheid, zodat het materiaal geschikt kon zijn voor vwo 3- en vwo 4-leerlingen. De twee toetsen samen bleken voldoende betrouwbaar voor groepsdifferentiatie (Cronbachs $\alpha = .66$). Afzonderlijk presteerden ze wisselend (cloze, met 20 items: Cronbachs $\alpha = .63$; tekst met vragen, met 9 items: Cronbachs $\alpha = .51$), waarbij de correlatie tussen beide opvallend gering bleek ($r = .06$). Kennelijk bevragen beide toetsen afzonderlijke dimensies van tekstbegrip.

Ten tweede construeerden we een zogenaamde *post-then-pre*-test. Daarin konden deelnemers na de lessenserie op een vijfpuntsschaal bij een zestiental stellingen aangeven hoe zij hun leesgedrag en tekstbegrip in het algemeen beoordeelden, vóór en ná de lessenserie. Een dergelijke test meet vooral de door deelnemers ervaren verandering door een interventie (Hill & Betz, 2005). Daarbij vroegen we leerlingen aan te geven hoe vaak ze bepaald leesgedrag inzetten zowel vóór als ná de lessenserie (Lam & Bengo, 2003).

We onderscheidde drie sets stellingen. Vijf stellingen betroffen de mate waarin leerlingen tekstelementen in eigen woorden omzetten tijdens lezen, conform het eerste doel van de lessenserie (diep begrip opbouwen); daarbij bevroeg de laatste van deze vijf de inschatting van leerlingen in hoeverre ze een tekst voldoende denken te begrijpen. Vijf stellingen betroffen de inzet van terugkijkgedrag, conform het tweede doel van de lessenserie (intensief teruglezen van tekstdelen); twee stellingen bevroegen hier aspecten die niet expliciet aan bod kwamen (terugkijken naar afzonderlijke / moeilijke woorden). Omdat we ook geïnteresseerd waren in eventuele neveneffecten, volgden zes stellingen over de inzet van andere leesstrategieën. Twee daarvan kwamen expliciet in de EMME-lessen aan bod, de andere impliciet. Zie voor de stellingen Tabel 5.3.

Betrokken docenten Nederlands corrigeerden de voor- en natoets. Een van de betrokken onderzoekers scoorde daarnaast afzonderlijk alle werken. Scores weken slechts incidenteel af. Afwijkingen werden in overleg in overeenstemming gebracht.

2.5 Data-analyse

Voor het meten van het effect van de interventie is gebruikgemaakt van multiveau-modellen, waarbij als vaste effecten conditie en leerjaar zijn meegenomen. Hierbij zijn de scores op de voormeting opgenomen als covariaat. We zijn dan geïnteresseerd in de verschillen in gemiddelde scores tussen condities, gegeven de scores op de voormeting. Bij het toetsen van de eventuele verschillen tussen condities in scores op de natoets en scores op de *post-then-pre*-test zijn observaties steeds genest binnen leerlingen.

3 Resultaten

3.1 Experimentafname en data-uitval

Het experiment werd afgenomen in de laatste maanden van 2020. De Covid19-epidemie in Nederland zorgde voor meer uitval in het experiment dan gebruikelijk. Dit gold vooral de natoetsen in de vwo 3-klassen die vrijwel onmiddellijk vóór de tweede lockdown in december 2020 werden afgenomen.

Tabel 5.1 *Uitval van participanten per conditie per leerjaar op experimentonderdelen (N = 131)*

Uitval van participanten	Voortoets	Natoets	Voor- óf natoets	Post-then-pre
Experimentele cond. (n = 90)	4	15	18	7
- waarvan vwo 3 (n = 50)	4	15	18	3
- waarvan vwo 4 (n = 40)	0	0	0	4
Controleconditie (n = 41)	0	13	13	5
- waarvan vwo 3 (n = 23)	0	13	13	5
- waarvan vwo 4 (n = 18)	0	0	0	0

In Tabel 5.1 is de uitval van het aantal leerlingen weergegeven voor de voortoets, de twee natoetsen en de *post-then-pre*-test. Omdat niet alle uitvallers op de voortoets ook uitvielen op de natoetsen, is in Tabel 5.1 ook weergegeven hoeveel leerlingen ofwel bij de voortoets ofwel bij de natoetsen afwezig waren. Voor de analyse van de voor- en natoetsen konden we dus 72 leerlingen uit de experimentele conditie en 28 leerlingen uit de controleconditie meenemen ($n = 100$). Voor de analyse van de *post-then-pre*-test konden we 83 leerlingen uit de experimentele conditie en 36 leerlingen uit de controleconditie meenemen. Uitgevallen leerlingen zijn overigens steeds uit de data weggelaten.

De lessenserie zelf verliep zonder verdere bijzonderheden, en de participatie was ruim voldoende tot zeer goed, blijktens het ingevulde lesmateriaal en naar de inschatting van de deelnemende docenten.

3.2 Voor- en natoetsen

De voortoets, die niet gepretest was, bleek tijdens de afname voldoende betrouwbaar voor groepsdifferentiatie (Cronbachs $\alpha = .63$).

De leerlingen in vwo 3 scoren op voor- en natoetsen afzonderlijk én tezamen niet aantoonbaar lager of hoger dan de leerlingen in vwo 4. In de analyse is daarom de factor leerjaar achterwege gelaten. Daarnaast: omdat de twee natoetsen elk andere aspecten van diep begrip bleken te bevragen, is in de analyse het effect van cloze-taak en tekst met vragen afzonderlijk meegenomen.

Tabel 5.2 Parameterschatting van de scores op beide onderdelen van de natoets

Parameter	Coëfficiënt	(se)	
<i>Fixed parameters</i>			
Cloze-natoets	9.58	(.49)	** ^a
ΔConditie Observerend lezen	1.36	(.58)	*
Vragen-natoets	3.94	(.30)	**
ΔConditie Observerend lezen	.79	(.37)	*
<i>Covariaten</i>			
Effect voortoets: cloze-natoets	.32 ^b	(.08)	**
Effect voortoets: vragen-natoets	.16 ^b	(.05)	**
<i>(Co)varianties tussen leerlingen</i>			
	<i>S² cloze</i>	<i>S² vragen</i>	
<i>S² cloze</i>	7.41 (1.00)	.26 ^c	
<i>S² vragen</i>	1.19 (.48)	2.86 (.39)	

^a significantie: * $p < .05$, ** $p < .01$; ^b gecentreerd rond gemiddelde; ^c correlatie tussen vragen- en cloze-natoets

Tabel 5.2 geeft een schatting van de scores op voor- en natoetsen tussen de condities, gecorrigeerd voor hun score op de voortoets. De score van een gemiddelde leerling op de cloze-natoets in de controleconditie is geschat als 7.59. Eenzelfde leerling heeft in de experimentele conditie een score van $(9.58 + 1.36 =) 10.94$. Per punt hoger (of lager) op de voortoets neemt zijn score met .32 toe (of af) op de cloze-natoets.

De score van een gemiddelde leerling op de vragen-natoets in de controleconditie is 3.94. In de experimentele conditie is de score van eenzelfde leerling hoger: $(3.94 + 0.79 =) 4.73$. Per punt hoger (of lager) op de voortoets neemt zijn score met .16 toe (of af) op de vragen-natoets.

3.3 Post-then-pre-test

Hoofdeffecten vanuit leerjaar en conditie zijn niet aantoonbaar: vwo 3-leerlingen schalen zichzelf in deze test niet aantoonbaar hoger of lager in dan vwo 4-leerlingen, noch schalen leerlingen in de controleconditie zichzelf over de gehele test genomen hoger of lager in dan leerlingen in de experimentele conditie.

Tabel 5.3 geeft voor elke stelling afzonderlijk een algemene schatting van de gemiddelde scores op de *post-then-pre-test*, zowel voor de leerlingen in de controleconditie als die in de experimentele conditie. De eerste set stellingen (stelling 1 tot en met 5) heeft betrekking op strategisch terugkijkgedrag. Hier geven de leerlingen in de experimentele conditie aan ná de lessenserie meer te zijn gaan terugkijken naar zinnen en alinea's, maar niet naar afzonderlijke woorden of moeilijke woorden, in vergelijking met de leerlingen in de controleconditie.

De tweede set stellingen (stelling 6 tot en met 11) heeft betrekking op meer indirect in de lessenserie besproken leesstrategieën. Hier geven de leerlingen in de experimentele conditie aan vaker (delen van) de tekst ná lezing te zijn gaan herlezen en meer op hoofdzaken te zijn gaan letten, zaken die ook expliciet in de experimentele lessenserie aan bod kwamen.

Tabel 5.3 Parameterschatting van de gemiddelde score per stelling op de post-then-pre-test (S^2_e variatie tussen leerlingen; S^2_r residuele variantie)

Stelling	Controle- conditie		Experimentele conditie			S^2_e ; S^2_r
	gem. vóór	gem. na	gem. vóór	gem. na		
1. lees ik woorden of woordgroepen terug.	3.37	3.49	3.16	3.51	^a	(.55; .25)
2. lees ik zinnen terug.	3.58	3.58	3.33	3.62	^{a c*}	(.64; .22)
3. lees ik een of meerdere alinea's terug.	3.19	3.33	2.57	3.12	^{a b c**}	(.73; .27)
4. stop ik bij een moeilijk woord dat ik niet ken.	3.19	3.30	3.49	3.60	^a	(.92; .17)
5. lees ik de tekst regel-voor-regel, in één keer door.	3.14	3.16	3.01	3.03		(.89; .24)
6. scan ik voor het lezen de tekst.	3.07	3.17	2.80	3.23	^{a c*}	(1.48; .20)
7. scan ik na het lezen de tekst.	2.65	2.74	2.99	3.38	^{a b c*}	(1.31; .18)
8. let ik tijdens lezen op de hoofdzaken in de tekst.	3.12	3.17	3.01	3.38	^{a c*}	(.84; .26)
9. pauzeer ik tijdens lezen om even na te denken.	3.14	3.18	2.41	2.62	^{a b}	(1.09; .17)
10. maak ik tijdens lezen aantekeningen, bijv. in de kantlijn.	1.58	1.72	1.46	1.83	^a	(.54; .35)
11. maak ik na lezen aantekeningen, bijv. in de kantlijn.	1.51	1.72	1.51	1.72	^a	(.30; .47)
12. formuleer ik hetgeen ik lees, in mijn hoofd om tot eigen woorden	3.14	3.14	2.78	3.19	^{a c**}	(.91; .21)
13. lees ik de tekst zo dat ik hem naderhand in mijn hoofd kan samenvatten.	2.72	2.81	3.10	3.65	^{a b c**}	(.91; .21)
14. kan ik na lezen zelfstandig de hoofdgedachte ervan formuleren.	2.88	3.02	2.91	3.35	^{a c**}	(.72; .13)
15. vraag ik me na lezen af of ik de tekst goed genoeg begrepen heb.	3.16	3.21	3.16	3.68	^{a c**}	(.91; .23)
16. denk ik dat ik na lezen de tekst inderdaad voldoende begrijp.	3.30	3.33	3.41	3.81	^{a b c**}	(.40; .20)

^a significant hoofdeffect voor/na de lessenserie; ^b significant hoofdeffect vanuit conditie; ^c significant interactie-effect voor/na lessenserie met conditie, waarbij * $p < .05$; ** $p < .01$

De derde set stellingen (12 tot en met 16) heeft betrekking op het opbouwen van diep begrip. Hier rapporteren de leerlingen in de experimentele conditie ook een grotere groei, in vergelijking met leerlingen in de controleconditie.

4 Discussie en conclusie

Met deze experimentele lessenserie wilden we aantonen dat observerend leren lezen, via EMME's, vwo-leerlingen bewust kan maken van wat diep tekstbegrip inhoudt, om zo bij te dragen aan diep begrip als zij zelf met teksten aan de slag gaan. Eerdere positieve resultaten van observerend leren lezen (Couzijn, 1995; Keehnen et al., 2015) konden we in deze studie bevestigen. Na slechts vijf lessen kon een significante toename in tekstbegrip bij vwo 3- en vwo 4-leerlingen worden aangetoond, zowel wanneer het tekstbegrip geëvalueerd werd met een cloze-toets als wanneer de evaluatie plaatsvond met tekstbegripsvragen.

Leerlingen in de experimentele conditie geven niet alleen bij de natoetsen blijk van meer tekstbegrip dan leerlingen in de controleconditie, ze lijken ook door de lessenserie een aanmerkelijk rijkere mentale representatie te hebben opgebouwd van wat diep begrip inhoudt. Leerlingen in de experimentele conditie geven aan dat ze ná de lessenserie vaker het gelezene in eigen woorden omzetten, vaker op hoofdzaken letten, vaker de hoofdgedachte formuleren, vaker alinea's herlezen en zich vaker afvragen of ze de tekst goed genoeg begrepen hebben, dan vóór de lessenserie. Deze rijkere mentale representatie lijkt dan ook een logische verklaring voor de hogere tekstbegripsscores op de natoets.

Bevreemdend is dit resultaat niet. Veel EMME-onderzoeken vinden positieve leereffecten (voor een overzicht: Van Marlen, 2019). Specifiek voor de leesles hebben EMME's op video's het voordeel dat leerlingen niet alleen het leesgedrag van helder en scherp reflecterende leeftijdsgenoten kunnen observeren en evalueren, maar ook van leeftijdsgenoten die dit beperkt of nauwelijks kunnen. Weliswaar waren deze *peer*-modellen in onze lessenreeks gescript, maar we hebben geen aanwijzing dat deelnemende leerlingen dit hebben waargenomen – in EMME's wordt overigens vaak video- of audiomateriaal gemanipuleerd (zie Van Marlen, 2019).

Daarmee vormen EMME's een vanzelfsprekend instrument om een klas strategisch te leren lezen. De laatste jaren is erop aangedrongen dat er in de leesles Nederlands meer wordt uitgegaan van een doelgerichte, flexibele benadering van evidencebased leesstrategieën (Van Steensel & Houtveen, 2020). Veel bestaande leergangen Nederlands behandelen in hun leesparagrafen echter geen leesstrategieën, maar een beperkte set 'generieke manieren van lezen' (hoofdstuk 1; zie ook Rooijackers et al., 2021b). Met deze EMME-lessenserie kan de effectiviteit van door modellen gebruikte strategieën voor het tekstbegrip klassikaal worden besproken en geëvalueerd, én kunnen evidencebased strategieën op relatief natuurlijke wijze in een klassengesprek aan bod komen. Natuurlijk kunnen docenten Nederlands een lessenserie als deze, met haar complexe, arbeidsintensieve constructieprocedure, niet eenvoudig zelf construeren. Te gelegener tijd hopen we deze lessenserie toegankelijk te maken voor docenten, waarmee ze als voorbeeld kan dienen voor, bijvoorbeeld, makers van leergangen Nederlands, die over meer tijd en middelen beschikken om lesmateriaal te construeren.

Ook een ander ontwerpprincipie achter deze lessenserie, naast observerend leren lezen, kan hebben bijgedragen aan de effectiviteit ervan. In de reguliere leesles Nederlands moeten leerlingen vaak de tekst vooraf zonder een duidelijk leesdoel bestuderen, waarna ze een ogenschijnlijk lukraak samengestelde set tekstafhankelijke vragen moeten beantwoorden. Dit stellen van vragen bij een tekst stimuleert hoogstwaarschijnlijk een diepere verwerking van de tekstinhoud (Rouet et al., 2001). Echter, doordat het voor leerlingen vooraf bij deze taak onduidelijk is waarop ze moeten letten, wordt het vooraf-lezen vooral een middel om de vragen correct te beantwoorden (zie hoofdstuk 4). Leerlingen lijken zo met deze taak een grote taakvaardigheid te ontwikkelen ('vraag 1 ging over alinea 1, dus vraag 2 zal wel over alinea 2 of 3 gaan': zie hoofdstuk 3). Doordat ze zich zo sterk op de vragen rich-

ten, lijken leerlingen ermee bovendien geen zelfstandige attitude tot diep begrip te ontwikkelen (zie hoofdstuk 4; vgl. Breukink, In voorbereiding).

Wij kozen daarom voor een andere taakinrichting. In de werklessen in de experimentele lessenserie formuleerden we voorafgaand aan de tekstbestudering steeds een duidelijk leesdoel en namen we, na initiële tekstbestudering, bij het uitdelen van de vervolgtask de tekst in. Als leerlingen vooraf weten op welke tekstdimensie de vervolgtask betrekking zal hebben én als ze vooraf weten dat de tekst niet langer raadpleegbaar is bij de vragen, worden vragen voor hen niet langer een doel op zich maar vooral een middel om te controleren in hoeverre ze vooraf voldoende tekstbegrip hebben opgebouwd; daarmee kunnen ze zich sterker richten op de vooraf-tekstbestudering en wordt de opbouw van diep begrip hoogstwaarschijnlijk bevorderd (vgl. Schroeder, 2011). Mits ingezet binnen een breed spectrum aan leestaken, kan een werkwijze als deze daarom een interessante alternatieve leestask zijn voor de leeslessen Nederlands.

Vooropgesteld, het betreft hier een quasi-experiment, afgenomen op één middelbare school, waarbij vooral korte uiteenzettende en opiniërende teksten werden gebruikt: terughoudendheid ten aanzien van de generaliseerbaarheid van deze studie is daarom gepast. Ook hebben we niet gemeten wat de retentie van de interventie is: in hoeverre het effect van de interventie enkele maanden na afname aantoonbaar is, weten we niet.

Al met al denken we echter te kunnen vaststellen dat leerlingen via EMME's op *peer*-niveau op diep tekstbegrip kunnen reflecteren, zodat ze scherper beseffen wat diep begrip bij een tekst inhoudt en vanuit dit scherpere besef vervolgens zelf teksten met meer begrip lezen. Gezien het belang dat de laatste jaren aan diep lezen en *close reading* voor het Nederlandse onderwijs wordt gehecht (Onderwijsraad & Raad voor Cultuur, 2019) én de recente zorgen over dalend tekstbegrip onder middelbare scholieren, kan een EMME-lessenserie als deze het vakdidactisch instrumentarium in de leeslessen Nederlands waardevol aanvullen.

HOOFDSTUK 6

De effectiviteit van een lessenserie
over diep lezen

*Een interventieonderzoek onder vwo 4-leerlingen**

* Dit hoofdstuk verschijnt in een (licht) gereviseerde vorm als: Rooijackers, P., Mantingh, E., Silfhout, G. van, & Bergh, H. van den. De effectiviteit van een lessenserie over diep lezen; een interventieonderzoek onder vwo 4-leerlingen. *Pedagogische Studiën*, 99(4).

Samenvatting

Men traint in Nederland tekstbegrip vaak op traditionele wijze: leerlingen bestuderen een tekst en beantwoorden vervolgens tekstafhankelijke vragen. Er loopt momenteel in Nederland een discussie of het tekstbegripsonderwijs op middelbare scholen niet inhoudelijk moet worden herzien; een alternatieve benadering van de tekstbegripsles lijkt wenselijk.

Gebaseerd op recente wetenschappelijke inzichten rondom tekstbegrip, is daarom een digitale lessenserie ontwikkeld die in 27 vwo 4-klassen ($N = 701$) is beproefd. In zes lessen staat 'diep lezen' van zakelijke teksten centraal; tekstbegrip wordt steeds functioneel ingebed en via observerend leren procesgericht benaderd, via een breed spectrum aan taken; het zelfstandig bestuderen van teksten staat voorop. De lessenserie werd afgenomen in twee condities: 'Zelfstandig leren' ($n = 329$), waarin leerlingen individueel de lessen doorwerkten, en 'Docent-gestuurd leren' ($n = 372$), waarin leerlingen de lessen deels via klassengesprekken doorwerkten.

Resultaten: in beide condities geven leerlingen zelf aan dat ze na de lessenserie meer leesgedrag zijn gaan inzetten dat voor diep lezen kenmerkend is; dit geldt sterker voor de conditie 'Docent-gestuurd leren'; in de conditie 'Docent-gestuurd leren' vertonen leerlingen na de lessenserie ook meer (diep) tekstbegrip dan vóór de lessenserie; in beide condities blijkt de kwaliteit van de wijze van uitvoering door leerlingen bepalend voor de leerwinst. Al met al lijkt de lessenserie daarmee een effectief eigentijds alternatief voor de nu gebruikelijke leesles Nederlands.

1 Inleiding

De laatste jaren is er internationaal meer aandacht gekomen voor de specifieke kenmerken van diep lezen (Allen & McNamara, 2020; Wolf, 2018). Diep lezen omvat “the array of sophisticated processes that propel comprehension and that include inferential and deductive reasoning, analogical skills, critical analysis, reflection, and insight” (Wolf & Barzillai, 2009, p. 32; vgl. Baron, 2015; Hayles, 2012; Piper, 2012; Wolf, 2018). Daarmee sluit diep lezen aan bij het theoretisch gefundeerde concept situatiemodel, zoals dat met name in de Constructie-Integratietheorie is uitgewerkt (Kintsch, 1998). Volgens Wolf (2018) zouden jonge lezers in de 21^e eeuw minder met diep lezen vertrouwd zijn dan voorgaande generaties; door de exponentiële informatietoename via internet en sociale media zouden jongeren juist bedreven zijn in scannend lezen, in *hyper-reading* (vgl. Liu, 2005; Sosnoski, 1999).

In Nederland, dat een sterk gedigitaliseerde samenleving kent (bijv. DESI, 2020), is deze analyse van de actuele leesvaardigheid van jongeren inmiddels door overheidsinstanties overgenomen: de jongeren in Nederland “lezen weliswaar volop korte tekstjes – berichten op hun smartphone of samenvattende stukjes in schoolboeken – maar besteden minder tijd aan ‘diep lezen’: het geconcentreerd lezen van langere teksten of boeken. Mede hierdoor gaat hun leesvaardigheid achteruit” (Onderwijsraad & Raad voor Cultuur, 2019, p. 2; vgl. Ministerie van Onderwijs, Cultuur en Wetenschappen, 2020; Pereira & Nicolaas, 2019).

1.1 Leesonderwijs in Nederland

Evenals in Nederlandstalig buurland Vlaanderen zijn er in Nederland de laatste jaren zorgen gerezen over het dalende tekstbegrip van scholieren. In de internationaal vergelijkende PISA-test vertonen Nederlandse vijftienjarigen sinds 2012 een dalende trend in hun tekstbegripsscore; in vergelijking met Europese leeftijdsgenoten blijken ze vooral minder sterk in evalueren en reflecteren (OECD, 2019). Deze laatste bevinding wordt weerspiegeld in de bekende PIRLS-test naar het tekstbegrip van basisschoolleerlingen: Nederlandse groep 6-leerlingen zijn aanmerkelijk sterker in het vinden van informatie dan in het reflecteren op de tekstinhoud (Gubbels et al., 2017). De zorgen hierover hebben inmiddels ook de politiek bereikt (Ministerie van Onderwijs, Cultuur en Wetenschappen, 2020). Het lezen van informatieve teksten is dan ook een van de meest basale vaardigheden in de huidige informatiesamenleving (Catrysse, 2017; OECD, 2019).

In Nederland wijkt de inrichting van het domein leesvaardigheid in enkele opzichten af van de internationale standaard. Waar bijvoorbeeld in leermiddelen in het Angelsaksische onderwijs het lezen van fictionele en narratieve teksten vaak nauwelijks tot niet wordt onderscheiden van dat van zakelijke teksten (bijv. Fisher & Frey, 2015), wordt dit onderscheid in Nederlandse leermiddelen en leeslessen vrijwel steeds aangehouden: het is terug te vinden in leergangen begrijpend lezen voor het basisonderwijs (bijv. Bogaerds-Hazenberg et al., 2022b), in leergangen van het schoolvak Nederlands voor het voortgezet onderwijs (bijv. Frank et al., 2019), in de door de overheid vastgelegde niveaubeschrijvingen voor taalvaardigheid in het Nederlandse primaire en secundaire onderwijs (Expertgroep Doorlopende Leerlijnen Taal, 2008), en in de centrale afsluitende toetsing van het voortgezet onderwijs (CvTE, 2022).

Er loopt momenteel in de Nederlandse politiek en het Nederlandse onderwijsveld een discussie of het taalonderwijs, en in het bijzonder het leesonderwijs in de Nederlandse taal, niet aan een fundamentele herziening toe is (bijv. Inspectie van het Onderwijs, 2019; Ministerie van Onderwijs, Cultuur & Wetenschappen, 2022; Onderwijsraad & Raad voor Cultuur, 2019; Pereira & Nicolaas, 2019). Vanuit de wetenschap zijn er eveneens zorgen geuit over het Nederlandse leesonderwijs (Van den Broek et al., 2021; Van Steensel et al., 2020). Op dit moment wordt er een grondige herziening van het schoolvak Nederlands voorbereid, die waarschijnlijk andere accenten en aanvullingen zal opleveren in de landelijke beschrijvingen van de beheersingsniveaus en examenprogramma's voor leesvaardigheid (vgl. Curriculum.nu, 2019; Rijksoverheid, 2022). We lopen een aantal van de gesignaleerde probleempunten in het huidige Nederlandse leesonderwijs langs.

Ten eerste: zoals wel in meer landen het geval is (bijv. Fang, 2016; Wijekumar et al., 2019), is in Nederland de taakinrichting van de lessen tekstbegrip vaak nogal eenzijdig: Nederlandse basisschoolleerlingen en middelbarescholieren lezen doorgaans een relatief lange tekst, zonder een functioneel leesdoel, en beantwoorden daarna, doorgaans individueel, een set tekstafhankelijke begripsvragen, waarna ze individueel, in groepjes of klassikaal hun antwoorden bespreken of nakijken (Bogaerds-Hazenberg et al., 2022b; Linthorst & De Glopper, 2015; SLO, 2021a). Al zijn er zeker positieve afwijkingen van deze standaardpraktijk (vgl. bijv. Pronk-Van Eunen & De Vos, 2018), daarmee spiegelt de Nederlandse praktijk in grote lijnen het leesonderwijs zoals dat in de klassieke studie van Durkin (1978) werd bekritiseerd: inzet van de leesles is nog altijd vaak of leerlingen een goed of fout antwoord op een vraag hebben. Op deze eenzijdige inrichting in de leesles is in Nederland al sinds decennia kritiek: leerlingen zouden ermee in hoge mate taakspecifieke verwachtingen en vaardigheden opbouwen, toetsing van tekstbegrip wordt verward met training, en leerlingen leren ermee vooral 'een kunstje' (bijv. Hoogeveen & Bonset, 1998; Hoogeveen, 2018; Westhoff, 2012). Voor deze kritiekpunten werd in eerder onderzoek onderbouwing gevonden: leerlingen lijken er inderdaad zeer specifieke taakverwachtingen mee op te bouwen (zie hoofdstuk 3) en ze leren met deze taak waarschijnlijk niet zelfstandig een tekst te bestuderen (zie hoofdstuk 4).

Een tweede probleempunt: in de les zakelijk lezen is er slechts in geringe mate aandacht voor het leesproces, en de aandacht die er is, kent duidelijke beperkingen.

In de leesles staan, zoals gezegd, doorgaans de leerlingantwoorden op begripsvragen centraal. Docenten richten zich met enige regelmaat op het leesproces via *modelling*, ze doen voor hoe ze zelf de inhoud van een (pittige) tekst zich eigen maken; hoe leerlingen zélf een tekst lezen en begrijpen, de effectiviteit van hun leesgedrag, blijft echter buiten beschouwing (Bogaerds-Hazenberg et al., 2020; Linthorst & De Glopper, 2015). Daarnaast is de vraag in hoeverre docenten in het Nederlandse voortgezet onderwijs in hun lessen evidencebased leesstrategieën (bijv. Duke et al., 2011) toepassen. Actuele leergangen Nederlands voor het voortgezet onderwijs bieden namelijk geen strategieën aan maar generieke manieren van lezen: oriënterend lezen, globaal lezen, studerend lezen, etc. (Schuppert & Van den Bergh, 2009; Rooijackers et al., 2021b). Deze methodiek is gebaseerd op verouderde inzichten uit de neerlandistiek, met name die in Drop en De Vries (1974), en verschilt mar- kant van internationale inzichten over effectieve strategie-instructie (bijv. Almasi & Fullerton, 2012).

Een derde probleempunt: de aandacht gaat in de leesles vooral uit naar enkelvoudig-tekstbegrip (*single text comprehension*) en de argumentatieve aanvaardbaarheid van informatie binnen een tekst (SLO, 2021a); meervoudig-tekstbegrip (*multiple text comprehension*) en de beoordeling van de kwaliteit, betrouwbaarheid en bruikbaarheid van informatie komen doorgaans weinig aan bod (Van den Broek et al., 2021; Gubbels, 2020; Rooijackers et al., 2021a). Deze meer 21e-eeuwse dimensies van leesvaardigheid hebben de laatste twee decennia internationaal ruim wetenschappelijke aandacht gekregen (bijv. Braasch et al., 2018; Britt et al., 2018) en spelen inmiddels een grote rol in de theoretische raamwerken van internationale toetsen als PISA en Pirls (OECD, 2019; Mullis et al., 2017), evenals in voor- stellen voor curriculumherzieningen in bijv. de VS (Pearson et al., 2020). Evalueren en reflecteren bij een of meer teksten “were arguably always part of reading literacy, but their importance has increased with the increased amount and heterogeneity of information readers are faced with today” (OECD, 2019, p. 36). In Nederland sluit de inhoud van leesparagrafen in leergangen, centrale toetsen en beschrijvingen van beheersingsniveaus daarentegen nog sterk aan bij inzichten uit *single-text*-leestheorieën (met name Kintsch, 1998). Er weerklinken in Nederland dan ook al geruime tijd geluiden deze meer eigentijdse begripsdimensies een prominentere plek in het onderwijs te geven (bijv. Gubbels, 2020; Kirschner, 2017; Rooijackers et al., 2021b; SLO, 2021b).

Een laatste probleempunt, samenhangend met het voorgaande: de inzet van tekstbegrip voor een schoolse taak, zoals bij een schrijfopdracht of presentatie, blijft in leergangen Nederlands nogal eens buiten beschouwing (vgl. Van den Broek et al., 2021; Pereira & Nicolaas, 2019; SLO, 2021a). Onderzoek laat echter zien dat vaardige lezers zich niet alleen van minder vaardige lezers onderscheiden in tekstbe- grip, maar ook in de wijze waarop ze dat begrip voor schoolse taken opbouwen en toepassen: leerlingen kunnen een opgegeven leestaak onderling bijvoorbeeld nogal afwijkend interpreteren en op grond daarvan afwijkende representaties van tek- sten opbouwen (Britt et al., 2018). Deze inbedding van tekstbegrip binnen een spe- cifieke, contextgebonden leestaak ontbreekt in de lessen Nederlands doorgaans.

Zonder een functionele, doelmatige context wordt lezen echter een geïsoleerde, schijnbaar 'doelloze' vaardigheid (bijv. Berenst et al., 2016; Van den Broek et al., 2021; Perreira & Nicolaas, 2019); bovendien leidt het integreren van leesvaardigheid in bijvoorbeeld schrijftaken vaak tot een beter begrip van teksten (bijv. Graham & Perin, 2007; Graham & Hebert, 2011). Er is in Nederland daarom wel voor gepleit om tekstbegrip consequent te koppelen aan kennisdomeinen in het schoolvak Nederlands (Meesterschapsteam Nederlands, 2019). Deze functionele inbedding van tekstbegrip Nederlands is dan ook onderwerp van discussie bij de aanstaande curriculumherziening van het schoolvak (Curriculum.nu, 2019; SLO, 2021b).

1.2 *Diep lezen centraal*

Het concept diep lezen, vaak beschreven als de tegenhanger van *hyper reading* (Sosnoski, 1999; Wolf, 2018; vgl. Baron, 2015; Hayles, 2012), vormt sinds het laatste decennium steeds vaker onderwerp van onderzoek (vgl. Allen & McNamara, 2020). Diep lezen sluit sterk aan bij het begrip situatiemodel zoals beschreven in de meest courante theoretische modellen over de constructie van tekstbegrip (vgl. McNamara & Magliano, 2009a), met name de Constructie-Integratie-Theorie (Kintsch & Van Dijk, 1978; Kintsch, 1998) en het Landschapsmodel (Van den Broek et al., 1999). In deze theorieën wordt een tekst beschouwd als een netwerk van hiërarchisch samenhangende, betekenisvolle beweringen, waarin lezers relaties leggen tussen woorden, zinnen en alinea's. Dit constructieproces mondt gaandeweg uit in een tekstbasis, een vrij letterlijke mentale representatie van de tekstinhoud, en in een situatiemodel, een mentaal model waarin lezers de tekstuele wereld integreren in hun eigen voorkennis en opvattingen. Pas op dit laatste niveau kennen lezers aan een tekst werkelijk begrip toe (Kintsch, 1998). Door de persoonlijke, actieve verwerking van de tekstinhoud, tijdens én na lezing, ontstaat bij lezers vervolgens diep begrip van een tekst, aldus Wolf (2018): doordat lezers een tekst geconcentreerd, herhaaldelijk lezen en zich op analyse, reflectie en evaluatie richten, zijn ze in staat de tekstinhoud uit te diepen: te analyseren en interpreteren, de kwaliteit en geloofwaardigheid ervan te bepalen, vanuit de tekstinhoud logische vervolgstappen te zetten, etc. (Wolf & Barzillai, 2009). Diep lezen is als concept sterk verbonden aan literatuurwetenschappelijke inzichten over de denkprocessen die met name literatuur van lezers kan vergen (vgl. Wolf, 2018; Hayles, 2012).

Deze geconcentreerde, convergente en herhaalde lezing van één tekst vormt in veel studies de tegenhanger van *hyper-reading* (vgl. Sosnoski, 1999; Wolf, 2018). Lezers blijken informatie op internet doorgaans eenmalig, non-lineair en scannend te lezen, vooral op zoek naar kernwoorden, springend van tekst naar tekst; het lineair, diep en langdurig geconcentreerd lezen van één tekst raakt daardoor op de achtergrond (bijv. Halamish & Elbaz, 2020; Liu, 1999; Mangen et al., 2013; Sosnoski, 1999; Wolf, 2018). Deze aangepaste leeswijze zouden lezers hebben ontwikkeld om de overdadige hoeveelheid beschikbare informatie op internet en social media efficiënt te kunnen verwerken (bijv. Baron, 2015; Piper, 2012; Wolf, 2018). Echter, met deze aangepaste leeswijze laten ze vaak tal van diepere begripsdimensies in een

tekst buiten beschouwing, zoals de betrouwbaarheid van de verstrekte informatie en de inhoudelijke consistentie ervan (Wolf, 2018). De dichotomie *hyper reading* en diep lezen wordt overigens nog weleens gerelativeerd: het is de vraag in hoeverre de moderne lezer door de hedendaagse overvloedige informatiestroom niet juist vooral nieuwe strategieën en nieuwe leeswijzen heeft ontwikkeld, waarbinnen *deep reading* een van de vele instrumenten is (Clowes, 2019; Jones & Hafner, 2012).

In hoeverre een lezer een tekst diep leest, is afhankelijk van leesdoel en leessituatie (Wolf, 2018). Een uitgebreid krantenartikel over een onderzoek kan een lezer thuis snel en *hyper-readend* tot zich nemen, maar hetzelfde krantenartikel kan een lezer diep lezen als het als basis dient voor een presentatie op werk of school. Sinds de millenniumwisseling is in internationaal leesonderzoek dan ook steeds meer deze intentionaliteit achter leesgedrag in kaart gebracht: de situationele en functionele inbedding van tekstbegrip (Britt et al., 2018; McCrudden et al., 2011; Van den Broek & Helder, 2017; etc.). Uiteenlopende leesdoelen en leessituaties blijken bij lezers te leiden tot deels afwijkende begripsprocessen en een afwijkende inzet van vaardigheden en strategieën: wie een tekst voor een schoolse opdracht bestudeert, zet andere vaardigheden en strategieën in tijdens het lezen en legt andere accenten in een tekst dan wie diezelfde tekst voor recreatieve doeleinden leest (Van den Broek & Helder, 2017; Yeari et al., 2015).

Deze intentionaliteit achter tekstbegrip wordt ook verdisconteerd in recente theoretische leesmodellen (Britt et al., 2018; vgl. Britt & Rouet, 2012; McCrudden & Schraw, 2007). Volgens het recente RESOLV-model, dat leesgedrag bij schoolse leestaken wil verklaren (Britt et al., 2018; zie ook Rouet et al., 2017), bouwen lezers bij een taak over een of meerdere teksten vooraf eerst een statisch contextmodel op. Daarin verdisconteren ze zaken als taakinterpretatie, de status van de ‘opdrachtgever’, hulpmiddelen en belemmeringen bij de taak, evenals een zelfbeoordeling bij deze taak. Vanuit dit contextmodel stellen lezers vervolgens een dynamisch taakmodel op: taakspecifieke doelen en scenario’s, waarin ze interpreteren wat ze moeten doen om de doelen te bereiken, en welke procedures ze daarbij moeten hanteren. Deze context- en taakmodellen sturen vervolgens de constructie van begrip bij een tekst of teksten bij een lezer. Vaardige lezers onderscheiden zich bij het opbouwen en toepassen van dergelijke context- en taakmodellen sterk van minder vaardige lezers (Britt & Rouet, 2012; Gil et al., 2010).

1.3 Strategie-instructie

Een belangrijk instrument om een tekst diep te lezen, vormen strategieën. Strategieën zijn voor een belangrijk deel persoons-, tekst- en doelafhankelijk (Almasi & Fullerton, 2012; Pressley & Afflerbach, 1995). Wanneer twee vaardige lezers eenzelfde tekst ‘strategisch’ lezen, blijken ze nogal andere accenten in een tekst te plaatsen en daarbij (deels) andere strategieën toe te passen (vgl. Guthrie et al., 1996; Pressley & Afflerbach, 1995). Desalniettemin zijn er duidelijke gemeenschappelijke kenmerken in het leesgedrag van vaardige lezers aan te wijzen, zoals het identificeren en herlezen van kernzinnen, het bepalen van de hoofdgedachte in een tekst,

het herlezen van ingewikkelde passages, etc. (vgl. Duke & Pearson, 2002; Duke et al., 2011; Pressley et al., 1995).

Een grondige herlezing van de belangrijkste elementen in een tekst lijkt een minimaal vereiste voor begrip bij een enigszins complexe tekst. Door zinnen, regels en alinea's terug te lezen, kunnen lezers zowel falend tekstbegrip repareren als hun begrip van een tekst controleren en verdiepen (Van den Broek & Helder, 2017). Dat geldt hoogstwaarschijnlijk *a fortiori* voor situaties waarin lezers een complexe tekst moeten begrijpen. Een belangrijke voorwaarde voor diep begrip is immers de onderdompeling van de lezer in de wereld van de tekst (Wolf, 2018). Het intensief teruglezen van tekstgedeelten is dan ook een overkoepelend kenmerk van veel van de expert-leesstrategieën (Pressley & Afflerbach, 1995).

Onderzoek naar de effectiviteit van strategie-instructie kent inmiddels een lange traditie (Duke & Pearson, 2002; Afflerbach et al., 2020). Een breed arsenaal aan onderzoek toont de effectiviteit van expliciete strategie-instructie aan (Donker et al., 2014; Graham & Harris, 1994; Hattie et al., 1996; Rogiers, 2019; etc.; vgl. ook Okkinga et al., 2018). Ouder onderzoek naar de effectiviteit van leesstrategie-instructie vertrekt nog weleens vanuit een min of meer vaste set leesstrategieën die onderwezen wordt (bijv. Oostdam & Bimmel, 1996; Palincsar & Brown, 1984; Lysynchuk et al., 1990; etc.); vanaf midden jaren negentig vertrekt interventieonderzoek naar strategie-instructie sterker vanuit de opvatting dat lezers zelfstandig en gecoördineerd een flexibel strategie-instrumentarium moeten kunnen toepassen in uiteenlopende leessituaties (bijv. Brown et al., 1996; Pressley et al., 1995). Onderzoek naar de situationele inbedding van tekstbegrip in de laatste twee decennia (zie 1.2) heeft de legitimatie van deze flexibele benadering alleen maar versterkt: niet alleen wisselen lezers in de wijze waarop ze begrip opbouwen, ze variëren ook sterk in de wijze waarop ze dit begrip in uiteenlopende leessituaties toepassen (Afflerbach et al., 2020). Goed strategieonderwijs zou daarom niet alleen begrip, maar ook het doel en de situationele inbedding van dit begrip moeten verdisconteren (ibidem; NAEP Reading Framework, 2017): met andere woorden, strategieonderwijs moet nadrukkelijk oog hebben voor de conditionele dimensie van strategiegebruik.

Bij strategieonderwijs kan *observational learning* (Bandura, 1986) een logisch vertrekpunt vormen. Bij dit leerprincipe voeren leerlingen niet direct zelf een taak uit maar observeren ze eerst de taakuitvoering van een model, reflecteren erop en evalueren het. Omdat leerlingen de taak niet zelf uitvoeren, hebben ze meer cognitieve ruimte beschikbaar om te leren: ze verzamelen zo metacognitieve en procedurele kennis over de taakuitvoering en kunnen deze vervolgens toepassen als ze zelf een soortgelijke taak gaan verrichten. Vooral binnen het domein schrijfvaardigheid is de positieve werking van observerend leren redelijk uitgebreid in kaart gebracht (Braaksma et al., 2004; Couzijn, 1995; Rijlaarsdam et al., 2008; etc.), maar ook binnen het domein leesvaardigheid zijn er inmiddels studies beschikbaar die de effectiviteit ervan laten zien. Leerlingen blijken na een lessenserie met *observational learning* vaker strategisch leesgedrag in te zetten dan leerlingen die de lessenserie niet gevolgd hebben (Salmerón et al., 2020; Salmerón & Llorens, 2019), en ze geven naderhand meestal ook van een beter tekstbegrip blijk (zie hoofdstuk 5;

Couzijn, 1995; Salmerón & Llorens, 2019). Gedragsveranderingen door observerend leren hoeven zich overigens niet steeds te vertalen in een meetbaar verhoogd resultaat op lees- of schrijfvaardigheidsproducten, aldus de *social learning theory* (Bandura, 1986; Bandura & Walters, 1977): iets leren over het proces uit zich niet steeds in een beter product. In Salmerón et al. (2020) bijvoorbeeld vertoonden de leerlingen in de eye-trackresultaten wel duidelijke gedragsveranderingen – maar dit leidde niet tot een aantoonbaar effect op hun tekstbegrip.

1.4 Doelen en onderzoeksvragen

Aansluitend op voornoemde ontwikkelingen in het Nederlandse (lees)onderwijs, construeerden wij een zelfstandig uitvoerbare, digitale module over diep lezen voor het voortgezet onderwijs, bestaande uit zes lessen. Met deze lessenserie stellen we onszelf twee algemene doelen: ten eerste willen we erin moderne inhoudelijke inzichten rondom tekstbegrip een ruime plaats geven, met name de intentionaliteit achter tekstbegrip (Britt et al., 2018; McCrudden et al., 2011), een brede conceptualisatie van evaluatie en beoordeling van teksten (OECD, 2019), en daarnaast inzichten rondom (de kenmerken van) diep begrip (Wolf, 2018).

Het tweede doel komt voort uit de uitzonderlijke educatieve omstandigheden, ontstaan door de Covid19-epidemie: we wilden de effectiviteit van de lessenserie laten zien bij een volledig zelfstandige, individuele afname ervan. Door de maandenlange nationale schoolsluiting gedurende twee lockdowns nam het digitaal lesgeven (blended learning en e-learning) in Nederland een hoge vlucht, met name in het voortgezet onderwijs (SLO, 2021a; Smeets, 2021). Doordat een heropening van de scholen onzeker leek, ontwierpen wij daarom de lessenserie als een vorm van e-learning, zodat deze, bij de leerling thuis of op school, als een individueel, zelfstandig maakbare lesmodule kon worden afgenomen.

Doordat de Nederlandse scholen bij de start van het schooljaar 2021-2022 wederom de deuren voor hun leerlingen openden, deed de gelegenheid zich alsnog voor om de ontwikkelde digitale lessenserie niet alleen zelfstandig, maar ook onder gedeeltelijke begeleiding van een docent af te nemen. De begeleidende rol van de docent vormt immers een cruciale meerwaarde voor leren (Hattie, 2012), een belang dat tijdens de lockdowns weleens werd onderschat (Smeets, 2021). Daardoor werd besloten het experiment in twee *experimentele condities* aan te bieden, naast een controleconditie: in de ene experimentele conditie werkten de leerlingen de lessenserie (vrijwel) geheel individueel en zelfstandig door, terwijl in de andere experimentele conditie de docent in drie van de zes lessen een sterk sturende rol had, met name door middel van klassengesprekken. Om de ecologische validiteit van deze studie te verhogen, hebben we ervoor gekozen docenten zelf het experiment af te laten nemen, zonder uitgebreide training of assistentie in de klas door onderzoekers.

De *wijze van uitvoering* van deze lessenserie door leerlingen is daarmee wel een nadrukkelijk aandachtspunt in deze studie. In een eerdere pilotstudie (zie hoofdstuk 5) werden de lessen in een klas in een mengvorm van individuele, groeps-

en klassikale lesactiviteiten afgenomen, door de docent Nederlands van de klas, samen met een onderzoeker. In deze pilotstudie bleken de leerlingen die een lessenserie van vier lessen hadden gevolgd, naderhand meer diep begrip te hebben opgebouwd dan leerlingen die deze lessenserie niet hadden gevolgd. De vraag is in hoeverre de resultaten uit deze studie in deze vervolgstudie gerepliceerd kunnen worden. Leerlingen werken het lesmateriaal immers (naargelang conditie: voor een belangrijk deel of geheel) zelfstandig en individueel door. e-Learning blijkt echter lang niet altijd een duidelijk leerrendement op te leveren. De interactie met de docent, die tussen leerlingen en die tussen leerlingen en het computersysteem hebben bijvoorbeeld een impact op de mate waarin leerlingen van een e-learning-module leren: zonder feedback van én interactie met anderen kunnen leerlingen een isolement ervaren dat hun engagement met de module in de weg staat (bijv. Bouhnik & Marcus, 2006; Liaw, 2008; Martínez et al., 2007; Udo et al., 2011). De vraag is dus in hoeverre leerlingen zich, zonder de begeleidende, bemoedigende en sturende rol van de docent én zonder de afwisseling in werkvormen en lesactiviteiten, voldoende engageren en passend op het lesmateriaal reageren, temeer daar de lessenserie in een aantal opzichten markant afwijkt van de gebruikelijke inrichting van de leesles Nederlands.

Met de lessenserie richten we ons specifiek op vwo 4-leerlingen. Er zijn wel zorgen geuit over het tekstbegrip van Nederlandstalige universitaire studenten (Van Houtven et al., 2010; Raad voor Nederlandse Taal en Letteren, 2015; Warps et al., 2021). Deze zouden over het algemeen te weinig kritisch kunnen lezen (Warps et al., 2021), terwijl juist academisch georiënteerde leerlingen later in studie en werk vaak geconcentreerd en diepgaand een tekst zullen moeten bestuderen en vanuit meerdere teksten over een onderwerp in kwaliteit wisselende informatie zullen moeten evalueren en integreren (Goldman, 1997; Hermida, 2009; Perfetti, 1997).

In deze studie onderzoeken we dus de effectiviteit van deze lessenserie over diep begrip van zakelijke teksten, onder vwo 4-leerlingen. We zullen daarbij enerzijds verkennen in hoeverre leerlingen na de lessenserie meer metacognitieve kennis over diep begrip hebben opgebouwd en anderzijds verkennen in hoeverre leerlingen ná de lessenserie blijken van meer diep begrip in begripstoetsen dan vóór de lessenserie: iets leren over het proces uit zich volgens de *observational learning theory* immers niet steeds in een beter product. Daarnaast onderzoeken we ook in hoeverre de wijze van uitvoering door leerlingen op de effectiviteit van de lessenserie van invloed is.

Daarmee staan in deze studie dus drie onderzoeksvragen centraal:

- a) In hoeverre maakt de lessenserie vwo 4-leerlingen bewust van (de voorwaarden voor) diep begrip bij zakelijke teksten, en wijkt dit af per conditie?
- b) In hoeverre geven de vwo 4-leerlingen die de lessenserie hebben doorlopen, blijk van meer diep begrip als ze daarna zelf zakelijke teksten lezen, en wijkt dit af per conditie?
- c) In hoeverre is de wijze van uitvoering (implementatie) van deze lessenserie van invloed op de effectiviteit ervan?

Vertrekkend vanuit de staande praktijk in de Nederlandse leesles, gebruikten we in de lessenserie geen narratieve of fictionele teksten – al staan deze natuurlijk bij uitstek diep lezen toe (vgl. Hayles, 2012; Wolf, 2018) – maar uiteenzettende en betogende zakelijke teksten uit kranten en (digitale) tijdschriften.

2 Methode

2.1 *Participanten*

Het experiment is afgenomen op zeven Nederlandse middelbare scholen, in het noorden, midden en zuiden van Nederland, in 27 vwo 4-klassen ($N = 701$). De experimentele lessenserie als geheel bestond uit zes lessen van ca. 40 minuten: drie observerend-leren-lessen, afgewisseld met drie werklessen met speciaal ontworpen leestaken. Alle lessen zijn gegeven onder toezicht van de betrokken docenten Nederlands ($N = 17$) van de school, zonder enige actieve inbreng van de onderzoekers.

Voorafgaand aan het experiment werd met de deelnemende docenten de randvoorwaarden voor deelname besproken. Hierna ontvingen de deelnemende docenten een uitgebreide handleiding bij het experiment, met daarin het volgende: een korte bespreking van de theoretische uitgangspunten achter de lessenserie, praktische informatie, uitgewerkte lesplannen en een technische handleiding voor het werken met de digitale app waarmee de lessenserie kon worden afgenomen. Docenten konden bij vragen, problemen of andere verlegenheidssituaties vóór en tijdens het experiment altijd contact opnemen met een van de onderzoekers.

2.2 *Experimentopzet*

Het experiment werd afgenomen in twee experimentele condities. Deze condities weken niet af in inhoud, maar enkel in instructiemodus. De eerste conditie, 'Zelfstandig leren', werd afgenomen op vier scholen, met dertien vwo 4-klassen ($n = 329$): in deze conditie werkten de leerlingen zelfstandig en individueel de zes lessen van de lessenserie door en had de docent geen noemenswaardige inhoudelijke inbreng. De tweede conditie, 'Docent-gestuurd leren', werd afgenomen op drie scholen, met veertien vwo 4-klassen ($n = 372$): in deze conditie ontvingen leerlingen klassikaal drie lessen (de observerend-leren-lessen – zie onder), waarin de docent les gaf en klassengesprekken leidde; de andere drie lessen (de werklessen) werkten de leerlingen zelfstandig en individueel door.

Om de effectiviteit van de lessenserie in het algemeen én het verschil in effectiviteit tussen de twee onderzoekscondities te kunnen meten, hebben we in beide condities gebruikgemaakt van een *switching-replications*-design (Shadish et al., 2002). Klassen zijn telkens per school gerandomiseerd verdeeld in twee groepen, die op wisselende momenten de interventie ontvangen. Beide groepen maakten allereerst een voortoets. Vervolgens ontving groep 1 als eerste de interventie, terwijl groep 2 regulier onderwijs volgde, onderwijs dat op een ander domein dan tekstbe-

grip betrekking had. Een tweede toets volgde. Minimaal drie weken en maximaal vijf weken nadat groep 1 met de interventie was gestart, ontving vervolgens ook groep 2 de interventie, terwijl nu groep 1 regulier onderwijs volgde. Een natoets mat ten slotte het tekstbegrip van beide groepen: deze werd afgenomen vrijwel onmiddellijk na het afronden van de lessenserie door groep 2 en drie à vier weken na het afronden van de lessenserie door groep 1. In Figuur 6.1 is dit design grafisch weergegeven voor beide condities.

Conditie Docent-gestuurd leren

Groep 1	T ₁	X	T ₂	T ₃
Groep 2	T ₁		T ₂	X
				T ₃

Conditie Zelfstandig leren

Groep 1	T ₁	X	T ₂	T ₃
Groep 2	T ₁		T ₂	X
				T ₃

Figuur 6.1 Grafische weergave van het switching-replications-design voor beide onderzoekscondities, waarbij T's de toetsmomenten aanduiden en de X'en afnamemomenten van het experiment in de twee groepen van elke conditie

Een *switching-replications-design* heeft *a priori* een aantal voordelen. Allereerst is er een belangrijk ethisch voordeel: door de interventie in tijd te spreiden, kunnen beide onderzoeksgroepen de interventie ontvangen. Daarnaast zijn er drie belangrijke methodologische voordelen: (a) groep 2 kan als controlegroep dienen terwijl groep 1 de interventie ontvangt; (b) beide groepen ontvangen de interventie, zodat het effect van de interventie tweemaal kan worden getoetst; (c) nadat ook groep 2 de interventie heeft gevolgd, kan de natoets van groep 1 als een *delayed retention test* gelden (Shadish et al., 2002). In onze studie wordt deze retentie overigens slechts binnen een beperkt tijdspad gemeten: al drie weken nadat groep 1 de lessenserie heeft afgenomen, kan de natoets zijn afgenomen.

Binnen deze studie, waarbij we gebruikmaken van twee condities, heeft dit design ten slotte het voordeel dat het ons toestaat niet alleen de effectiviteit binnen een conditie te onderzoeken, maar ook het eventuele verschil in effectiviteit tussen beide onderzoekscondities.

2.3 Materiaal

Algemene opzet lessenserie. Het centrale leerdoel achter de lessenserie was dat we leerlingen kennis wilden laten opbouwen over diep begrip en dat we ze wilden leren doelgericht, bewust en zelf-regulerend (strategisch) tot diep begrip van een tekst te komen (Afflerbach et al., 2008; Almasi & Fullerton, 2012). We boden in de lessenserie daarom geen vaste set leesstrategieën aan, maar wilden leerlingen vooral laten

reflecteren op de effectiviteit van hun eigen leesgedrag. Observerend leren (Bandura, 1986) vormde daarbij een schragend leerprincipe: de lessenserie wisselde steeds af tussen een observerend-leren-les en een werkles. In de observerend-leren-les observeren leerlingen twee *peer*-modellen die duidelijk variëren in het tekstbegrip dat ze opbouwen. Leerlingen reflecteren op het opgebouwde begrip van de twee modellen en op de leesmiddelen die deze hierbij inzetten, én leerlingen evalueren dit (vgl. Braaksma, 2002). In de werklessen kunnen leerlingen daarna de (meta) cognitieve kennis uit de observerend-leren-les zelf op leestaken toepassen.

Daarbij wilden we dat leerlingen doelbewust en zelfstandig de in de lessenserie opgenomen teksten zouden lezen, met als doel diep begrip voor een specifieke leestaak. Dat diepe leesgedrag probeerden we op twee manieren af te dwingen. Eerste manier: de leestaken werden in de werklessen zo geconstrueerd dat leerlingen vooraf de tekst intensief moesten bestuderen, waarna ze bij de eigenlijke verwerkingstaak (bijv. begripsvragen beantwoorden, een sorteertaak maken) niet langer over de tekst zouden beschikken. In eerder onderzoek naar het lezen van teksten en het beantwoorden van begripsvragen (zie hoofdstuk 3) bleken vwo-leerlingen bij het vooraf-lezen van de teksten vaak geneigd tot oppervlakkig leesgedrag, waarna ze zich vooral bij het beantwoorden van de begripsvragen zetten tot een intensievere tekstraadpleging. Het feit dat leerlingen weten dat ze tijdens de verwerkingstaak de tekst nog kunnen raadplegen, lijkt een belangrijke verklaring voor dit oppervlakkige vooraf-leesgedrag (vgl. hoofdstuk 4). Wanneer lezers weten dat ze bij het beantwoorden van begripsvragen juist niet meer de tekst kunnen raadplegen, zetten ze zich vaak wél tot zo'n diepgaande vooraf-bestudering: lezers blijken de tekst dan vooraf langer te bestuderen, met meer oog voor diepe betekenis, en hun antwoordscore op begripsvragen hangt dan vooral samen met de mate waarin ze tijdens vooraf-lezen een adequate coherente mentale representatie van de tekstinhoud hebben gemaakt (Schroeder, 2011; vgl. Ferrer et al., 2017; Higgs et al., 2017; Ozuru et al., 2007). Omdat we wilden dat leerlingen zelfstandig zo'n coherente mentale representatie van de tekstinhoud opbouwden, benadrukten we dus dat leerlingen de verwerkingstaak moesten kunnen maken op basis van hun vooraf-bestudering van de tekst – en dat ze daarom niet langer de tekst bij de verwerkingstaak konden raadplegen.¹

Tweede manier: we wisselden bij elke werkles sterk in de vorm van de uiteindelijke leestaak. Door veelvuldige blootstelling aan één taak (de tekst met vragen) lijken Nederlandse leerlingen in de leesles zich vooral te richten op de taak, namelijk het

¹ Men zou de vraag kunnen stellen in hoeverre het valide is dat leerlingen bij de verwerkingstaak niet langer de beschikking hebben over de bestudeerde tekst, zoals in onze studie het geval is. In veel school-, studie- of werksituaties hebben lezers immers juist wél de beschikking over een tekst, wanneer ze om informatie verlegen zitten – hoewel zeker niet in alle situaties: denk aan discussies en presentaties, en leertoetsen, waarin leerlingen doorgaans informatie hebben gememoriseerd. Deze vraag lijkt ons binnen onze studie echter niet relevant: de aandacht gaat in deze studie uit naar de zelfstandige diepe bestudering van teksten, vanuit een vooraf gesteld, specifiek leesdoel. De verwerkingstaak ná de bestudering is in de lessenserie vooral een controlemiddel voor leerlingen om te bezien in hoeverre ze de tekst tijdens deze bestudering voldoende hebben doorgrond – waardoor de verhouding tussen tekstbestudering en verwerkingstaak anders ligt dan gebruikelijk.

beantwoorden van vragen: een zelfstandige bestudering van een tekst schiet er dan gemakkelijk bij in (zie hoofdstuk 3; Bogaerds-Hazenberg et al., 2022b). Daarom gebruikten we in de lessenserie naast tekstafhankelijke vragen (Fisher & Frey, 2015), ook sorteertaken (McNamara & Kintsch, 1996) en samenvattings- en argumentatietaken (Gil et al., 2010).

Daarnaast wilden we leerlingen via de lessenserie drie typen kennis over diep lezen laten opbouwen. Effectieve, kwaliteitsrijke strategie-instructie heeft idealiter op drie typen kennis betrekking: declaratieve, procedurele en conditionele kennis (bijv. Lorch et al., 1993; Paris et al., 1983). Vooral de conditionele inbedding van strategieën en vaardigheden ontbreekt in het Nederlandse leesonderwijs nogal eens (Bogaerds-Hazenberg et al., 2022b; Rooijackers et al., 2021b). Daarom werd de lessenserie opgedeeld in drie sets van twee lessen, waarbij elke set vooral een van deze drie kennistypen centraal stelde.

De conditionele inbedding van begrip vertaalde zich ook in twee andere ontwerpprincipes. Ten eerste werden alle leestaken in de lessenserie binnen een functionele context aangeboden (McCrudden et al., 2011). Leestaken werden steeds voorafgegaan door een voor leerlingen herkenbaar scenario, met een duidelijk leesdoel, waarvoor diep lezen noodzakelijk was. Men denke aan een voordracht over een artikel, een voorbereiding via een artikel op schrijftoetsen, deelname aan een discussie naar aanleiding van een artikel, etc. Ten tweede wilden we leerlingen niet alleen diep begrip laten opbouwen en/of hierop reflecteren, maar we wilden ze ook confronteren met de vraag hoe ze een adequaat taakmodel bij een leestaak opbouwen: bouwen modellen, en leerlingen zelf, wel het juiste begrip *voor de taak* op? We gebruikten daarom (in de laatste observerend-leren-les) een tekst die ogenschijnlijk vooral niet bruikbaar leek voor een opdracht en (in de laatste werkles) een tekst die juist ogenschijnlijk zeer bruikbaar leek, maar, vanuit diep begrip, dat nauwelijks of niet bleek te zijn. Daarbij werd in deze twee lessen aanvullend meervoudig-tekstbegrip (*multiple text comprehension*) ingezet, zodat de bruikbaarheid van een besproken tekst beter kon worden bepaald.

Tabel 6.1 *Overzicht van leerdoelen en leeractiviteiten in lessenserie met bijbehorende lesnummers*

Leerdoel	Les	Leeractiviteit				
		observerend leren		individueel werken	voortgang monitoren	
		via EMME	via animatie	zelfstandig	quiz-vragen	open vragen
declaratieve kennis van diep lezen	1-6	1, 3, 5	-	2	2	2
procedurele kennis van diep lezen	3-6	3, 5	-	2, 4	4	4
conditionele kennis van diep lezen	5-6	-	5	2, 4, 6	6	6
reflectie op taakmodel bij leesopdracht	5-6	-	5	6	5, 6	6
zelfstandig begrip opbouwen	1-6	-	-	2, 4, 6	-	-
scenario-based taak	1-6	-	-	-	-	-
meervoudig-tekst-begrip	6	-	-	6	-	-

Ten slotte: we hanteerden het principe van de *gradual release of responsibility* (Fisher & Frey, 2013). Omdat we met name de gemiddeld vaardige en minder vaardige vwo 4-lezer met deze lessenserie wilden bedienen – lezers die waarschijnlijk minder geverseerd zijn in diep begrip –, werden in de eerste lessen teksten gekozen en opdrachten geconstrueerd met een gemiddelde moeilijkheidsgraad. In de laatste twee lessen werd de complexiteit opgevoerd en werd het concept diep begrip rijk uitgewerkt. In Tabel 6.1 is voor elk van de zes lessen de leerdoelen en -activiteiten aangegeven.

Materiaal experimentele lessenserie: de eerste twee observerend-leren-lessen. In les 1 en 3, de eerste twee observerend-leren-lessen, reflecteerden leerlingen via modellen op de vragen: (1) ‘Wat is diep begrip?’ (declaratieve kennis) en (2) ‘Hoe bouw je het op?’ (procedurele kennis). Omdat begripsconstructie door lezers doorgaans een vrij fijnmazig proces is, gebruikten we voor deze modellen EMME’s: *Eye-Movements Modeling Examples* (vgl. Jarodzka et al., 2013). Een proefpersoon heeft eerst een taak uitgevoerd op een beeldscherm, terwijl een eye-tracker diens oogbewegingen (fixaties) op het beeldscherm registreerde. Een onderzoeker construeert vervolgens van deze oogbewegingen op het scherm een filmpje en monteert het audiocommentaar van de proefpersoon eronder. Zo’n EMME kunnen docenten vervolgens in de les gebruiken om met leerlingen op leeractiviteiten te reflecteren. Een EMME biedt duidelijke didactische voordelen: leerlingen krijgen een vrij directe inkijk in de complexe perceptuele en cognitieve processen van een lezer; bovendien blijkt het bekijken van EMME’s vaak motiverend (Salmerón & Llorens, 2019). Veel internationale didactische studies vonden dan ook positieve leereffecten bij EMME-inzet in de klas (zie voor een overzicht: Van Marlen, 2019).

Als basis voor de EMME’s maakten we gebruik van de eye-track-data uit een eerder *mixed-method*-onderzoek (zie hoofdstuk 3). In elke les boden we via EMME’s twee modellen aan die sterk contrasteerden in leesvaardigheid (Salmerón & Llorens, 2019). De vaardige lezers in de filmpjes onderscheidden zich duidelijk van de minder vaardige lezers a) in leestempo, want vaardige lezers zijn doorgaans snellere lezers dan minder vaardige lezers (Perfetti & Stafura, 2014); en b) in strategiegebruik, want vaardige lezers zetten doorgaans vaker en efficiënter strategieën in die het opgebouwde begrip bij een tekst verdiepen (Van den Broek & Helder, 2017). Per model werden kenmerkende fragmenten gebruikt van in totaal doorgaans een anderhalve minuut.

In het audiocommentaar reflecteerden de model-leerlingen achteraf op de eye-trackbeelden van hun leesgedrag: waarom keken ze terug naar eerdere zinnen of tekstdelen? Begrepen ze naar hun eigen idee wat ze lazen? De modellen varieerden hier sterk in a) de mate waarin ze in eigen woorden de tekstinhoud al dan niet adequaat weergaven, en b) de kwaliteit waarmee ze reflecteerden: vaardige lezers kunnen namelijk doorgaans scherper het rationale achter eigen leesgedrag verwoorden dan minder vaardige lezers (McNamara & Magliano, 2009b).

Bij de leerlingtaken voor deze twee EMME-lessen analyseerden leerlingen niet alleen de EMME’s afzonderlijk, maar moesten ze ook beargumenteren welk model op basis van de vertoonde beelden en het hoorbare commentaar naar hun idee

het duidelijkst bij de gelezen tekst diep begrip opbouwde, en welk model het minst (vgl. Braaksma, 2002). Tijdens verschillende stadia van de constructie werd dit materiaal gescreend door een groep screeners en er vonden meerdere try-outs plaats.

Materiaal experimentele lessenserie: de derde observerend-leren-les. Les 5, de derde observerend-leren-les richtte zich op de vraag ‘Wanneer en waarom bouw je diep begrip op?’ (conditionele kennis). Omdat in deze les vooral de kwaliteit van het taakmodel bij een leestaak centraal stond (Britt et al., 2018), gebruikten we voor de modellen in deze les geen EMME’s, maar video-animaties waarin de modellen hardop nadachten tijdens het verrichten van een complexe leestaak.

Hiertoe ontwierpen we allereerst een leestaak waarbij leerlingen een korte tekst zich eigen moesten maken voor een beleidswerkgroep voor een schoolkantine. Een zestal vwo 4-leerlingen verrichtte daarna hardop de taak, terwijl we audio-opnames van hun hardop-denkproces maakten. Hieruit werden twee opnames geselecteerd die contrasteerden in de kwaliteit en scherppte a) waarmee een leerling de inhoud van de tekst kon weergeven, b) waarmee deze reflecteerde op de bruikbaarheid van de tekst, en c) waarmee deze de tekstuele gedachtegang toepaste binnen het scenario.

In de animaties stond het audiocommentaar van de modellen centraal, terwijl screenshots van alinea’s in de tekst werden getoond waarop het commentaar betrekking had. Leerlingen hoorden alleen het audiocommentaar tijdens de belangrijkste stappen in het denkproces van de modellen. De animaties werden zo gecomprimeerd van ca. 10 minuten naar een duur van 2-3 minuten.

Constructie materiaal experimentele lessenserie: de werklessen. Voor les 2, 4 en 6, de drie werklessen, zijn vier korte uiteenzettende of betogende teksten uit Nederlandse dagbladen of tijdschriften geselecteerd. Bij elke leestaak is een scenario ontwikkeld waarvoor diep begrip noodzakelijk was. Daarna werd daarbij een passende verwerkingstaak ontwikkeld. Bij elke werkles werd sterk gewisseld in de vorm van de uiteindelijke leestaak.

Elke tekst werd voorafgegaan door de instructie dat de tekst ná bestudering, bij het specificeren van de verwerkingstaak, niet meer mocht worden ingezien, zodat leerlingen aan de hand van de verwerkingstaak konden controleren in hoeverre ze de tekst vooraf voldoende diep hadden gelezen. In Bijlage 3 is een voorbeeldtaak uit een van de werklessen uit de experimentele lessenserie opgenomen.

Digitale implementatie van de lessenserie. Voor de digitale afname van de lessenserie werd een Nederlandse online onderwijsapp gebruikt (LessonUp, lessonup.com). Deze werkwijze had als belangrijk voordeel dat alle antwoorden van leerlingen centraal konden worden vastgelegd en inzichtelijk waren voor de onderzoekers. Tijdens het afnemen van de lessenserie bood de app daarnaast docenten de mogelijkheid de voortgang van hun leerlingen te volgen, de antwoorden van leerlingen in te zien en deze eventueel van feedback te voorzien.

Tijdens de verwerking van de inhoud van de lessenserie in het format van de app bleek dat een van de ontwerpprincipes van de lessenserie technisch niet realiseer-

baar was: leerlingen kregen in de lessenserie steeds de nadrukkelijke instructie dat ze een tekst na bestudering niet meer mochten inzien bij de verwerkingstaak; in de uiteindelijke versie was dat niet het geval. Om te controleren of leerlingen zich aan onze instructie hielden, voegden we in twee lessen hierover een controlevraag in, na het afronden van de leestaak.

Onderscheid tussen condities. Om het onderscheid tussen condities te kunnen waarborgen, werden er twee varianten van de lessenserie ontwikkeld. Als eerste werd het materiaal voor de conditie ‘Zelfstandig leren’ geconstrueerd, waarbij er zorg voor werd gedragen dat de leerdoelen en instructies helder en volledig waren, taken zelfstandig maakbaar waren én dat er na elke leestaak duidelijke feedback werd gegeven.

Het materiaal voor de conditie ‘Docent-gestuurd leren’ werd hiervan afgeleid. De drie werklessen werden in hun geheel ongewijzigd overgenomen. De drie observerend-leren-lessen daarentegen werden voor klassikale afname aangepast: de uitleg bij de leerdoelen en de instructie bij taken werden ingekort, individuele taken werden vrijwel steeds vervangen door een sheet die met een prikkelende vraag uitnodigde tot een klassengesprek, en de feedback op antwoorden was beknopt, om bespreking ervan door de docent mogelijk te maken.

2.4 Controle-instrumenten implementatie

Om meer zicht te krijgen op hoe de lessenserie per conditie in de les werd geïmplementeerd, ontwikkelden we een aantal controle-instrumenten, waarbij we gebruikmaakten van een *mixed-methods*-aanpak: we verzamelden parallel aan elkaar tijdens en na afname van het experiment zowel kwantitatieve als kwalitatieve data (Creswell & Plano Clark, 2018). Na het verzamelen van de data werden ze afzonderlijk van elkaar gecodeerd en geanalyseerd, om de interpretatie van de data aan te kunnen scherpen.

Logboeken docenten. Deelnemende docenten ontvingen een logboek waarin hen werd verzocht notities te maken over het verloop van de lessenserie en van de afname van de toetsen. Na de les moesten ze aangeven of leerlingen de les volledig hadden kunnen afronden, of leerlingen geconcentreerd en gemotiveerd werkten, in hoeverre leerlingen gemakkelijk dan wel moeilijk vonden en welke lesactiviteiten de docent zelf had verricht. In het logboek voor de conditie ‘Docent-gestuurd leren’ werd aanvullend gevraagd of de leerlingparticipatie voldoende was tijdens de klassengesprekken; in de conditie ‘Zelfstandig leren’ werd gevraagd of leerlingen (zoals geïnstrueerd) zelfstandig en individueel werkten.

Interviews docenten. Na afloop van het experiment werd per school de coördinerende docent Nederlands ondervraagd door een van de onderzoekers, via een voor het experiment ontwikkeld semi-gestructureerd interview. Vooraf was de coördinerende docent gevraagd de indrukken rondom het experiment bij deelnemende docenten Nederlands te inventariseren.

Doel van het interview was tweeledig: ten eerste om te controleren in hoeverre de experimentele lessenserie was geïmplementeerd zoals beoogd; ten tweede om te achterhalen hoe docenten de lessenserie inhoudelijk beoordeelden.

Speciaal punt van aandacht bij de conditie 'Docent-gestuurd leren' was de vraag in hoeverre docenten het voeren van klassengesprekken tijdens de observerend-leren-lessen geslaagd vonden: in hoeverre slaagden ze erin met de leerlingen adequaat op het leesgedrag van de modellen te reflecteren? In hoeverre lukte het de modellen onderling te vergelijken? In hoeverre gaven de leerlingen de gewenste respons?

Punt van aandacht bij de conditie 'Zelfstandig leren' was de vraag in hoeverre leerlingen gedurende de hele lessenserie individueel en zelfstandig en daarbij geconcentreerd en gemotiveerd werkten.

Lesobservaties. Om te controleren op het onderscheid in lesactiviteit tussen de twee condities, maakten we gebruik van lesobservaties. Per conditie observeerden we op één school telkens de lesactiviteiten van twee klassen en hun docenten; per klas werden in totaal vier lessen geobserveerd, waarvan steeds twee observerend-leren-lessen en twee werklessen.

Voor de lesobservatie werd gekozen voor een werkwijze geïnspireerd op die van Elving-Heida (2019). Daarbij werden zowel de docent als drie leerlingen geobserveerd via een schema op basis van systematische directe observatie (vgl. Briesch et al., 2010). Bij de constructie van het observatieschema werden enkele uitgangspunten uit het State-Event Classroom Observation System toegepast (SECOS: Saudargas, 1997).

Voor de docentobservatie gold de volgende handelwijze: de onderzoeker observeerde de docent per minuut telkens tien seconden en legde diens leshandeling vast gedurende dit korte tijdvak. De lijst mogelijke handelingen omvatte enerzijds klassikale activiteiten - instructie geven, met de klas bronnen bekijken, klassengesprekken voeren - en anderzijds individuele activiteiten: een leerling hulp bieden. Handelingen waarbij de docent zich niet actief met de klas bezighield - zoals het observeren van leerlingen, het voorbereiden van lesactiviteiten, het corrigeren van leerlingwerken -, werden eveneens genoteerd.

Een vergelijkbare handelwijze werd gevolgd bij het observeren van de leerlingen. Eerst werden drie leerlingen gekozen die goed zichtbaar waren voor de onderzoeker; geselecteerde leerlingen werden niet geïnformeerd dat ze werden geobserveerd. Vervolgens observeerde de onderzoeker elke leerling per minuut telkens tien seconden en legde met een observatieschema de lesactiviteit van de leerling vast gedurende dit tijdvak. Het observatieschema omvatte enerzijds aan de lesinhoud gerelateerde activiteiten - schoolwerk, overleggen met docent - en anderzijds niet aan de lesinhoud gerelateerde activiteiten - om zich heen kijken, kletsen met klasgenoot en andere dingen doen.

Lesmateriaal. Aan de hand van de digitale app was eenvoudig meetbaar in hoeverre leerlingen ook daadwerkelijk alle onderdelen van de lessenserie hadden door-

gewerkt. De app registreerde per les a) het percentage dia's van een les die een leerling had bekeken, b) het percentage gecontroleerde (goede) antwoorden op de vragen in een les: leerlingen hadden bij elke gesloten vraag de mogelijkheid meerdere malen antwoord te geven, totdat ze het goede antwoord hadden.

Om te kunnen achterhalen of leerlingen zich aan de instructie hielden om *niet* terug te kijken naar een bestudeerde tekst bij de verwerkingstaak, stelden we bij twee werklessen aan het einde van de lessenserie de vraag of leerlingen zich aan deze instructie hadden gehouden.

Evaluatie lessenserie. Om te kunnen bezien hoe leerlingen de lessenserie evalueerden, voegden we aan het einde van de lessenserie enkele evaluerende vragen toe: we vroegen aan leerlingen aan te geven in hoeverre ze de lessenserie als geheel leerzaam en op niveau vonden en we vroegen om eventuele suggesties voor verbetering.

2.5 Meetinstrumenten effect lessenserie

Post-then-pre-test. Om te bezien in hoeverre leerlingen ander strategisch leesgedrag vertoonden ná de lessenserie, construeerden we een zogenaamde *post-then-pre-test*. Daarin konden leerlingen na de lessenserie op een vijfpuntsschaal bij een twintigtal stellingen aangeven hoe zij zelf hun leesgedrag en tekstbegrip in het algemeen beoordeelden, vóór en ná de lessenserie. Een dergelijke test meet vooral de door deelnemers ervaren verandering door een interventie (Hill & Betz, 2005). Daarbij vroegen we leerlingen aan te geven hoe vaak ze bepaald leesgedrag inzetten zowel vóór als ná de lessenserie (Lam & Bengo, 2003): bijvoorbeeld 'Als ik een tekst diep moet lezen, dan breng ik in mezelf de belangrijkste denkstappen van de tekst in kaart.'

We onderscheidden vier sets stellingen – zie voor alle stellingen de Tabel in Bijlage 4. Vijf stellingen betroffen de inzet van terugkijkgedrag, conform het uitgangspunt van flexibele strategie-inzet (intensief teruglezen van tekstdelen); twee stellingen bevroegen hier aspecten die niet expliciet aan bod kwamen (terugkijken naar afzonderlijke / moeilijke woorden). Vijf stellingen betroffen de mate waarin leerlingen tekstelementen in eigen woorden omzetten tijdens lezen, een basisvoorwaarde voor het bereiken van diep begrip; daarbij bevroeg de laatste van deze vijf de inschatting van leerlingen in hoeverre ze een tekst voldoende denken te begrijpen. Vijf stellingen betroffen dimensies van diep begrip: de betrouwbaarheid en bruikbaarheid bepalen van bronnen, vervolgdienstapen zetten vanuit de tekstinhoud. Omdat we ook geïnteresseerd waren in eventuele neveneffecten, waren vijf stellingen toegevoegd over de inzet van andere leesstrategieën. Twee daarvan kwamen expliciet in de EMME-lessen aan bod, de andere impliciet.

Voor-, tussen- en natoetsen. Om te kunnen controleren op verschillen in tekstbegrip vóór en na de interventie bij beide experimentele groepen, gebruikten we een voortoets, tussentoets en twee natoetsen. Als voor-, tussen- en (eerste) natoets werd

een cloze-toets gebruikt, een gatentekst, waarin woorden strategisch zijn weggelaten (de ‘*rational deletion*’-variant: vgl. Cain et al., 2005; Oller & Jonz, 1994; Gellert & Elbro, 2013; Kleijn, 2018). Er loopt sinds de jaren tachtig van de vorige eeuw een discussie over de vraag in hoeverre gatenteksten tekstbegrip breder kunnen meten dan op zinsniveau. Voor de ‘traditionele’ vorm van de gatentekst, waarin woorden in een vaste interval (elk N^e woord) uit de tekst worden weggelaten, lijkt deze kritiek deels valide: slechts een minderheid van de items doet in studies dan aantoonbaar een appel op begripsvaardigheden over zinsgrenzen heen (voor deze discussie: Pearson & Hamm, 2005; zie ook Kleijn, 2018). Voor de ‘*rational deletion*-variant’ daarentegen, waarin onderzoekers een strategische keuze maken in de weggelaten woorden (bijv. connectieven, modale uitdrukkingen en bepalingen, verwijzingen en antecedenten), lijkt deze kritiek niet op te gaan: in deze vorm weerspiegelt de testscore vooral de vaardigheid van lezers om op macroniveau coherentie in een tekst aan te brengen (vgl. Cain et al., 2005; Gellert & Elbro, 2013; Kleijn, 2018; Oller & Jonz, 1994). De cloze-test heeft als belangrijke voordelen dat ermee een relatief groot aantal items binnen kort tijdsbestek kan worden afgenomen, ze doorgaans eenduidig te beoordelen is en ze tekstbegrip vaak relatief betrouwbaar meet (ibidem).

Als tweede natoets gebruikten we een reguliere tekst-met-vragen-test, met zowel open als gesloten vragen. Wij gebruiken deze tweede natoets vooral, omdat de begripsvragentest in het schoolvak Nederlands zeer gebruikelijk is, al staat deze situatie de laatste tien jaar in Nederland scherp ter discussie (vgl. bijv. Hoogeveen, 2018; Kamalski, 2007). In hoeverre, in welke mate en onder welke condities een begripsvragentest valide tekstbegrip kan meten, lijkt af te hangen van een complexe set interacterende factoren van tekst- en vraagkenmerken, waardoor begripsvragentesten zeer wisselende begripsconstructen lijken te bevragen (bijv. Cutting & Scarborough, 2006; Keenan et al., 2008). Omdat we het effect van onze lessenserie ook wilden meten via een in het onderwijs vertrouwde toetsvorm - gatenteksten worden in het schoolvak Nederlands immers weinig toegepast -, hebben we desalniettemin deze toets als tweede natoets opgenomen.

De vier gebruikte toetsen werden eerder gepretest en bleken toen afdoende betrouwbaar (zie hoofdstuk 5). Deelnemende docenten Nederlands corrigeerden de drie cloze-toetsen. Een van de betrokken onderzoekers corrigeerde ter controle daarnaast per school de toetsen van één klas. De overeenstemming in correctie tussen deze twee beoordelaars bleek bij deze steekproef bij de drie gatenteksten (voor- en tussentoets, eerste deel natoets) steeds zeer hoog ($\alpha \geq .97$), waarschijnlijk doordat na uitgebreid pretesten een zeer nauwsluitend correctiemodel was geconstrueerd.

Wat de natoets met tekstafhankelijke begripsvragen betreft: deze werd beoordeeld door een bovenbouwdocent Nederlands en een van de onderzoekers. Deze scoorden onafhankelijk van elkaar alle natoetsen. Ook hier bleek de overeenstemming in beoordeling hoog ($\alpha = .95$). Afwijkingen in score tussen deze twee beoordelaars werden in onderling overleg in overeenstemming gebracht.

2.6 Data-analyse

Pre-then-post-test. Voor het meten van het effect van de interventie op deze toets is gebruikgemaakt van multiniveau-modellen, waarbij als vaste effecten conditie en de interactie tussen conditie en interventie is meegenomen. Bij het toetsen van de eventuele verschillen in scores op de post-then-pre-test zijn observaties steeds genest binnen leerlingen en klassen.

Voor-, tussen- en natoetsen. Voor het meten van het effect van de interventie is hier eveneens gebruikgemaakt van multiniveau-modellen, waarbij observaties steeds genest zijn binnen leerlingen en binnen klassen. Daarbij vertrekken we vanuit een vaste modellenopbouw: als vaste effecten gelden a) tijdstip van meting (de vier afgenomen toetsen), b) interventie, c) tijdstip van interventie (*switching replications*) en d) conditie. Daarnaast nemen we e) de interactie tussen conditie en interventie mee: we zijn dan geïnteresseerd in de verschillen in gemiddelde scores tussen condities, veroorzaakt door de interventie.

Uitgaande van deze vaste modellenopbouw, controleren we vervolgens in hoeverre de twee implementatievariabelen een effect hebben op de tekstbegripscore, zowel als interactie-effect met interventie als interactie-effect met interventie en conditie. Nu voegen we telkens een effect toe, zodat we van elk toegevoegd model de *fit* en passingsverbetering kunnen controleren. In Bijlage 5 is de modellenopbouw in een Tabel veraanschouwelijkt.

Een belangrijk voordeel van het werken met multiniveau-modellen is dat er mee alle beschikbare data kan worden meegenomen in de parameterschattingen; deels ontbrekende data bij een proefpersoon leiden niet ertoe dat alle data van deze proefpersoon buiten beschouwing moeten worden gelaten. Dat is in onze studie extra van belang: doordat afname tijdens de Covid19-epidemie plaatsvond, vielen er immers meer ontbrekende data bij leerlingen te verwachten in het experiment dan onder normale omstandigheden.

3 Resultaten

3.1 Implementatie van de lessenserie

Om zicht te krijgen op de implementatie van de lessenserie, worden de logboekgegevens, interviews, lesobservaties en de gegevens uit de digitale app in onderstaande getrianguleerd. Eerst wordt de deelname van leerlingen aan de lessenserie langsgelopen, vervolgens het onderscheid tussen condities, en ten slotte worden enkele implementatievariabelen nader bekeken.

Participatie lessenserie. Het experiment werd afgenomen in de eerste maanden van het schooljaar 2021-2022. De Covid19-epidemie in Nederland zorgde voor meer lesuitval in het experiment dan gebruikelijk en leidde ertoe dat enkele docenten

gaandeweg het experiment gedwongen waren een enkele (werk)les als huiswerk op te geven.

Volgens de logboeken van de docenten werd de complete lessenserie door vrijwel alle 27 klassen doorlopen. Drie klassen vormden daarop volgens de logboeken een uitzondering: door ziekte van één docent maakte één klas de twee laatste lessen niet; door onvoorziene lesuitval en bijzonder verlof van een docent maakten twee klassen op één school de twee laatste lessen als huiswerk, zonder dat de docent hierop later controleerde. Vier klassen maakten één werkles als huiswerk, één klas maakte twee werklessen als huiswerk – de docenten van deze klassen controleerden wel dit huiswerk.

Vanuit de digitale data, beschikbaar via de app, zien we dit beeld bevestigd. Per les namen gemiddeld 630 van de 701 leerlingen deel. Vooral bij de eerste vier lessen is de aanwezigheid relatief hoog (per les: $gem. = 646$, $sd = 14$), terwijl deze bij de twee laatste lessen aanmerkelijk lager ligt (per les: $gem. = 598$, $sd = 2.5$).

Bij aanwezigheid blijken leerlingen de dia's van een les doorgaans vrijwel in hun geheel door te werken.² Voor de drie werklessen, voor de twee condities samen, geldt dat een leerling bij deelname gemiddeld 97.5% ($sd = 10.9$) bekeek van de dia's van de lessen. Dit percentage verschilt tussen condities ($F(2, 634.34) = 5.40$; $p = .02$), al is dit verschil beperkt ('Docent-gestuurd leren': $gem. = 97\%$, $sd = 11.8$; 'Zelfstandig leren': $gem. = 98.2\%$, $sd = 9.7$). Dit verschil ontstond vooral doordat docenten in de conditie 'Docent-gestuurd' lessen vaak kort klassikaal openden en daarmee de eerste dia's van een les bespraken. Voor de drie observerend-leren-lessen vinden we eveneens hoge percentages, die daarbij wel duidelijk verschillen per conditie ($F(2, 73571.10) = 133.47$; $p < .001$): in de conditie 'Zelfstandig leren' bekijkt een leerling bij aanwezigheid gemiddeld 94.3% ($sd = 20.3$) van de dia's van een les door. Voor de conditie 'Docent-gestuurd leren' is dat percentage aanmerkelijk lager: gemiddeld 81.5% ($sd = 26.6$). Dit grote verschil is eenvoudig verklaarbaar: de docent gaf in deze conditie immers vooral klassikaal les: een leerling kon de voortgang van de les zowel volgen op het eigen scherm als op het digitale klassenbord.

Overigens dient hier gesignaleerd dat sommige docenten via de digitale app erop controleerden dat leerlingen de gehele les doorwerkten. Op twee scholen in de conditie 'Zelfstandig leren' werd de werkwijze gehanteerd dat een leerling in principe elke les volledig doorwerkte en bij absentie de gemiste lessen inhaalde.

Onderscheid tussen condities: docentactiviteiten. Bekijken we allereerst de lesactiviteiten van de docenten. Bij de drie *observerend-leren-lessen* geven docenten uit de conditie 'docent-gestuurd leren' in logboeken en interviews steeds aan een groot deel van hun tijd te hebben besteed aan instructie en klassengesprekken, in tegenstelling tot docenten uit de conditie 'Zelfstandig leren', die aangeven hieraan geen tijd te hebben besteed.

² Voor de goede orde: we nemen hier aan dat leerlingen een dia hebben bekeken, als ze deze minimaal vier seconden op het scherm zichtbaar hebben gehad. In hoeverre leerlingen dan met aandacht deze dia's hebben bekeken, valt daarmee evenwel niet te zeggen.

In de acht geobserveerde lessen vinden we deze bevindingen bevestigd. De hoeveelheid geregistreerde instructietijd van docenten in een les verschilt van conditie tot conditie ($F(2, 861.13) = 15.53; p < .01$): in de conditie 'Zelfstandig leren' is de hoeveelheid instructietijd 11% ($sd = 9.27$), terwijl deze in de conditie 'Docent-gestuurd leren' 31.8% ($sd = 5$) is. Ook de hoeveelheid tijd besteed aan het klassikaal bekijken van bronnen wijkt af ($F(2, 496.13) = 90.90; p < .001$): in de conditie 'Zelfstandig leren' beslaat dit 1.3% van de lestijd ($sd = 1.5$), terwijl in de conditie 'Docent-gestuurd leren' dit 17% is ($sd = 2.9$). Er is daarnaast een zeer duidelijk verschil aantoonbaar in de hoeveelheid bestede tijd aan klassengesprekken ($F(2, 2278.13) = 475.44; p < .001$): deze vinden in de conditie 'Zelfstandig leren' in het geheel niet plaats, terwijl het percentage in de andere conditie 33.8% ($sd = 3.1$) is.

Wat de drie *werklessen* betreft: docenten geven in logboeken en interviews aan zich hier niet actief opgesteld te hebben: ze voerden geen klassengesprekken, noch gaven ze uitgebreide instructies, zoals de bedoeling was. Een tweetal docenten gaf in de interviews aan dat de hoeveelheid hulpvragen van individuele leerlingen tijdens de les wel sterk kon variëren. Deze opmerkingen stroken met de bevindingen in de acht geobserveerde lessen: docenten voerden geen klassengesprekken, noch bekeken ze met de klas bronnen.

Alle resultaten wijzen erop dat docenten zich dus inderdaad aan de lesinstructie in beide condities hebben gehouden.

Onderscheid tussen condities: leerlingactiviteiten. Bekijken we dan de handelingen van leerlingen. Docenten geven in logboeken en interviews aan dat leerlingen doorgaans het grootste deel van hun tijd besteden aan de lessenserie. Wel gaven enkele docenten aan dat het orde houden de nodige tijd vergde. In vier klassen werkten leerlingen daardoor regelmatig samen of was de concentratie laag.

Binnen de acht geobserveerde lessen is er een duidelijk verschil tussen condities waar te nemen. Zo is er allereerst een verschil aantoonbaar tussen de hoeveelheid tijd die de leerling gemiddeld zichtbaar aan de lesinhoud besteedt ('schoolse taak') ($F(2, 936.33) = 26.35; p < .001$): leerlingen in de conditie 'Docent-gestuurd leren' besteden gemiddeld 87.8% ($sd = 7.6$) hieraan, terwijl leerlingen in de conditie 'Zelfstandig leren' 96.7% ($sd = 3.8$) hieraan besteden. De tijd dat een leerling in (klassikaal) gesprek met een docent is, varieert eveneens per conditie ($F(2, 14.08) = 11.60; p = .001$): leerlingen in de conditie 'Docent-gestuurd leren' besteden gemiddeld 4.1% ($sd = 1.6$) hieraan, terwijl leerlingen in de conditie 'Zelfstandig leren' geen tijd hieraan besteden. Ten slotte is er ook een verschil in tijd dat een leerling met een medeleerling praat ($F(2, 500.52) = 31.42; p < .001$): leerlingen in de conditie 'Docent-gestuurd leren' besteden gemiddeld 8.8% ($sd = 4.9$) hieraan, terwijl leerlingen in de conditie 'Zelfstandig leren' 2.4% ($sd = 2.9$) hieraan besteden. Waarschijnlijk doordat docenten in de eerste conditie veelvuldig met klassen in gesprek gingen, praatten leerlingen vaker met elkaar, al dan niet over de lesinhoud.

Leerlingen besteden dus in de conditie 'Zelfstandig leren' hun lestijd grotendeels aan het doorwerken van de lessenserie. In de conditie 'Docent-gestuurd'

leren is er meer interactie tussen docent en leerlingen én meer interactie tussen leerlingen onderling.

Implementatie. Om te bezien in hoeverre leerlingen zich aan de instructies in de lessenserie hielden en zich engageerden met de lessenserie, bouwden we enkele controle-instrumenten in.

Ten eerste werd bekeken in hoeverre leerlingen bij gesloten vragen in de lessen hun *antwoord controleerden*. Omdat er in de conditie ‘Docent-gestuurd leren’ in de drie observerend-leren-lessen nauwelijks gesloten vragen aanwezig waren, is het zinnig onderscheid te maken tussen werk- en observerend-leren-lessen. Bekeken over de drie werklessen, is het percentage gecontroleerd goede antwoorden hoog (87.9%, $sd = 11.94$). Dit percentage wijkt duidelijk af tussen condities ($F(2, 3348.60) = 24.53; p < .001$). In de conditie ‘Docent-gestuurd leren’ is dit percentage 85.4% ($sd = 13.5$), terwijl het in de conditie ‘Zelfstandig leren’ hoger ligt: 90.4% ($sd = 9.5$). In de drie observerend-leren-lessen in de conditie ‘Zelfstandig leren’ ligt het percentage eveneens hoog: 95.4% ($sd = 9.9$).

Ten tweede werd leerlingen twee maal aan het einde van een werkles gevraagd of ze, zoals geïnstrueerd, *niet terug hadden gekeken* naar de tekst bij de vervolgoopdracht. Bij de eerste werkles gaf 27% van de leerlingen aan zich niet aan deze instructie te hebben gehouden; bij de tweede werkles gaf 28% dit aan – percentages die overigens niet per conditie verschilden ($p \geq .51$). Het aandeel leerlingen dat aangeeft terug te kijken, overlapt tussen beide lessen slechts gedeeltelijk ($r = .39$): 38.5% van de leerlingen blijkt hiermee zich minimaal één keer niet aan de instructie te hebben gehouden.

In de logboeken en interviews gaven docenten in de conditie ‘Docent-gestuurd leren’ aan dat ze soms moeite hadden om een goed lopend klassengesprek te voeren over de geobserveerde modellen, omdat zowel de docenten als de leerlingen zelf nogal onervaren waren met een dergelijke procesgerichte benadering. Ook wezen vier docenten uit deze conditie erop dat leerlingen niet altijd het gedrag van de modellen doorgrondde: “leerlingen vinden het redelijk gemakkelijk, maar uit de klassengesprekken blijkt juist dat ze echt nog niet alles doorzien.”

Samenvattend. Ondanks het feit dat het experiment tijdens de Covid19-epidemie werd uitgevoerd, was de deelname door leerlingen ruim voldoende en verliep de afname zonder noemenswaardige incidenten. Leerlingen die aan een les deelnamen, lijken doorgaans ook de lesinhoud vrijwel geheel te hebben doorgewerkt.

Vanuit de logboeken, interviews en lesobservaties over docent- en leerlinggedrag rijst het beeld op dat docenten en leerlingen grotendeels conform de opgestelde afnamecondities hebben gehandeld; in de conditie ‘Zelfstandig leren’ lijken docenten (nauwelijks tot) niet actief te hebben deelgenomen aan de lessen en werkten leerlingen veelal zelfstandig; docenten in de conditie ‘Docent-gestuurd leren’ gaven vaak klassikale instructie, bekeken vaak met leerlingen bronnen en voerden vaak klassikale gesprekken. De instructies in de lessenserie zelf lijken leerlingen redelijk te hebben opgevolgd: ze controleerden over het algemeen zeer vaak

hun antwoorden in de observatie- en werklessen. Echter, aan de instructie om bij de verwerkingsopdracht *niet* terug te kijken naar de bestudeerde tekst, hielden lang niet alle leerlingen zich: 38.5% van de leerlingen gaf aan hiervan te zijn afgeweken. Ook gaven sommige docenten uit de conditie ‘Docent-gestuurd leren’ aan dat ze het moeilijk vonden goede klassengesprekken te voeren over de leerlingmodellen in de EMME’s.

3.2 *Evaluatie van de lessenserie*

Evaluatie door leerlingen. Wanneer leerlingen zelf gevraagd wordt de lessenserie te evalueren na de laatste les, geeft 83.5% aan dat ze de lessenserie ‘goed op niveau’ vindt; 10.3% vindt de lessenserie ‘gemakkelijk’; 6.3% vindt de lessenserie ‘moeilijk’. 86.4% van de leerlingen geeft na de lessenserie aan deze leerzaam te hebben gevonden.

Op het verzoek om suggesties te geven voor verbetering van de lessenserie gingen leerlingen vaak niet in. Er werden vooral korte complimenten uitgedeeld (“erg tof”, “leuk om te doen”). Sommige leerlingen gaven aan dat ze de lessenserie wat te lang vonden en adviseerden inkorting. Het vaakst geopperde advies was dat één type vraag in de lessenserie (zogenaamde ‘sleepvragen’) technisch anders zou moeten worden ingericht.

Docenten gaven aan dat leerlingen over het algemeen redelijk enthousiast waren over de lessenserie, al vonden ze de lessenserie wel vaak aan de lange kant. Ook zouden leerlingen de lessenserie over het algemeen vrij gemakkelijk hebben gevonden. Verder gaven docenten van drie scholen aan dat leerlingen moeite hadden met de inhoudelijke opzet van de lessenserie. Ze misten ‘leerbare inhoud’, zoals lijsten met functietermen en argumentatieschema’s en vonden het accent op de opbouw van begrip verwarrend. Twee scholen boden de leerlingen daarom voorafgaande aan de natoetsen pdf’jes aan met de dia’s van drie van de zes lessen uit de lessenserie.

Evaluatie door docenten. Docenten evalueerden de lessenserie wisselend. De docenten van vijf van de zeven scholen waren redelijk tot zeer enthousiast: ze gaven aan de gekozen benadering ‘verfrissend’ te vinden (“een eye-opener”) en het idee te hebben dat ze leerlingen nieuwe inzichten in tekstbegrip konden bieden (“Het helpt leerlingen om zicht te krijgen in het leesproces en dat te spiegelen aan wat ze zelf kunnen” en: “ik vind het heel mooi om te zien hoe ze in zes lessen opbouwen naar vragen over het ontkrachten van argumenten en afleiden van tegenargumenten; (...) op zo’n niveau zijn we in klas 4 eerder niet aan de slag geweest.”)

Vooraf op één school waren er zeer duidelijke twijfels over de lessenserie. De twee docenten van deze school hadden twijfels over de moeilijkheid van de lessenserie (“het niveau is voor vwo te laag”) en gaven aan dat ze niet inzagen wat de meerwaarde van de lessenserie vormde (“geen goede voorbereiding op het Centraal Examen Nederlands”) en dat leerlingen naar hun idee vooral plichtmatig in de lessenserie hadden geparticipeerd.

3.3 Effect van de lessenserie

Ervaren leereffect: post-then-pre-test. 600 leerlingen (86% van de deelnemers) maakten deze test. De uitval op deze test in het experiment was daarmee relatief hoog, vooral doordat twee docenten elk in één klas de test niet bleken afgenomen te hebben. Elke set van vijf samenhangende stellingen bleek afzonderlijk ruim voldoende betrouwbaar ($GLB \geq .76$). Alleen de set stellingen die niet expliciet besproken strategieën betrof, scoorde duidelijk lager in betrouwbaarheid, zij het nog altijd voldoende (vóór de lessenserie: $GLB = .65$; na de lessenserie: $GLB = .73$).

Tabel 6.2 Parameterschatting van de gemiddelde score op de post-then-pre-test per set stellingen, per conditie (S^2_k : variantie tussen klassen; S^2_l : variantie tussen leerlingen; S^2_r : residuele variantie)

Set stellingen	Docent-gestuurd leren		Zelfstandig leren			$S^2_k; S^2_l; S^2_r$
	vóór (gem./se)	na (gem./se)	vóór (gem./se)	na (gem./se)		
1. strategisch terugkijken	2.76 (.04)	3.46 (.04)	2.77 (.04)	3.28 (.04)	^{a,c}	(.00; .32; .71)
2. opbouwen van begrip	2.73 (.04)	3.47 (.04)	2.79 (.05)	3.30 (.05)	^{a,c}	(.01; .31; .86)
3. dimensies van diep begrip	2.14 (.05)	2.83 (.05)	2.28 (.06)	2.84 (.06)	^{a,c}	(.01; .37; .91)
4. ander strategisch leesgedrag	2.23 (.05)	2.92 (.05)	2.24 (.06)	2.78 (.06)	^{a,c}	(.01; .29; 1.17)

^a significant hoofdeffect voor/na de lessenserie; ^b significant hoofdeffect conditie;

^c significant interactie-effect voor/na lessenserie met conditie

In Tabel 6.2 zijn voor elke set stellingen de schattingen vóór en na de lessenserie per conditie weergegeven. (In de Tabel in Bijlage 4 zijn deze schattingen per stelling te vinden.) In beide condities geven leerlingen voor alle sets stellingen aan dat ze ná de lessenserie vaker diep-leesgedrag zijn gaan vertonen, dan dat ze dat vóór de lessenserie vertoonden. Leerlingen in de conditie 'Docent-gestuurd' rapporteren daarbij een duidelijk grotere gedragswijziging dan leerlingen in de conditie 'Zelfstandig leren'.

Leerlingen geven dus over het algemeen aan dat ze na de lessenserie vaker strategisch terug zijn gaan kijken, vaker (diep) begrip zijn gaan opbouwen en vaker andere strategieën zijn gaan inzetten, dan dat ze dat vóór de lessenserie deden. Leerlingen in de conditie 'Docent-gestuurd leren' rapporteren hierbij steeds een grotere gedragsverandering dan de leerlingen in de conditie 'Zelfstandig leren'.

Effect op tekstbegrip van de lessenserie. De voortoets, die uit 27 items bestond, bleek bij het experiment voldoende betrouwbaar om te kunnen differentiëren tussen groepen ($N = 668$; $GLB = .76$), evenals de tussentoets, die uit 23 items bestond ($N = 628$; $GLB = .74$). De twee natoetsen ($N = 656$) presteerden wisselend: de gatentekst, die uit 20 items bestond, bleek voldoende betrouwbaar ($GLB = .72$), de tekst met vragen, die uit 9 items bestond, was dat duidelijk niet ($GLB = .48$) – een probleem dat door het gebruik van twee natoetsen grotendeels wordt opgevangen. In Tabel 6.3 zijn de correlaties tussen de vier toetsen weergegeven: we zien dan dat de correlaties tussen de drie cloze-toetsen ($r \geq .34$) steeds hoger liggen dan de correlatie van de drie cloze-toetsen met de begripsvragentoets ($r \leq .32$).

Tabel 6.3 *Correlaties (Pearson's r) tussen de vier gebruikte toetsen in het experiment*

correlatie	voortoets	tussentoets	natoets: cloze	natoets: vragen
voortoets	-	.38	.44	.26
tussentoets	.38	-	.34	.25
natoets: cloze	.44	.34	-	.32

In Bijlage 5 is de *fit* en passingsvergelijking van elk toegevoegd model nader beschreven; hier rapporteren we enkel in hoeverre een toegevoegd effect, per toegevoegd model, significantie bereikt. Uitgaande van de vooraf opgestelde modellenopbouw, vinden we allereerst een hoofdeffect voor tijdstip van meting ($F(3, 702.46) = 812.17; p < .001$), wat gezien de afwijkende maximumscore tussen de gebruikte toetsen een evidente uitkomst is. We vinden evenwel geen hoofdeffect voor interventie ($F(1, 407.86) = .05; p = .82$), noch vinden we een onderscheid tussen de twee momenten (*switching replications*) waarop de interventie plaatsvond ($F(1, 132.04) = .87; p = .35$). We vinden evenmin een hoofdeffect voor conditie ($F(1, 55.27) = 1.26; p = .27$). Een interactie-effect tussen conditie en interventie is wel aantoonbaar ($F(1, 919.02) = 6.46; p = .01$): de tekstbegripsscore wijkt na de interventie af tussen beide condities.

Nemen we vervolgens de twee *implementatievariabelen* van de lessenserie mee. Kijken we allereerst naar de eerste controlevariabele: het percentage gecontroleerde antwoorden tijdens de drie werklessen. Deze blijkt als interactie-effect met interventie aanwezig ($F(1, 545.53) = 4.03; p = .045$). Een interactie-effect tussen interventie, conditie en deze implementatievariabele is echter niet aantoonbaar ($F(1, 598.82) = .19; p = .67$).

Kijken we dan naar de tweede controlevariabele: de vraag of leerlingen zich aan de instructie hielden om niet terug te kijken naar de tekst bij de vervolgpdracht. Deze blijkt noch aantoonbaar als interactie-effect met interventie ($F(1, 524.29) = 1.12; p = .29$), noch aantoonbaar als interactie-effect met interventie en conditie ($F(1, 523.54) = .45; p = .50$).

In Tabel 6.4 zijn de parameterschattingen gegeven voor de scores op de vier toetsen in het experiment. Daarin is de eerste implementatievariabele als interactie-effect met interventie verdisconteerd, die immers een aantoonbaar was; de tweede implementatievariabele, die geen aantoonbaar interactie-effect met interventie kende, is daarin buiten beschouwing gelaten. We zien dan dat de leerlingen in de conditie 'Docent-gestuurd leren' na de interventie gemiddeld hoger scoorden op de toetsen dan dat ze dat vóór de interventie deden ($\beta = .66, se = .24$); het betreft hier overigens een klein effect (Cohen's $d = .22$). Voor de conditie 'Zelfstandig leren' kunnen we een dergelijk verschil niet aantonen ($\beta = -.40, se = .31$). Dit effect per conditie blijkt niet afhankelijk van het tijdstip van de interventie: leerlingen die op het eerste tijdstip de interventie ontvingen, scoorden niet aantoonbaar hoger of lager dan de leerlingen in de tweede conditie ($\beta = .20, se = .21; p = .34$).

Daarnaast: naarmate leerlingen hun antwoorden percentueel vaker controleeren, scoorden ze onafhankelijk van conditie gemiddeld hoger op de toetsen ná de interventie, dan dat ze dat vóór de interventie deden ($\beta = .016, se = .01; p = .045$).

Tabel 6.4 Parameterschattingen voor de score op de vier toetsen in de lessenserie

Parameter	coëfficiënt	(se)	p
<i>Meetmomenten</i>			
voortoets (conditie 'Zelfstandig')	17.10	(.32)	<.001
tussentoets	13.91	(.38)	<.001
natoets: cloze	10.60	(.42)	<.001
natoets: vragen	5.96	(.41)	<.001
Δ conditie 'Docent-gestuurd'	-.37	(.32)	.26
Δ tijdstip interventie	.20	(.21)	.34
Δ interventie : conditie 'Zelfstandig'	-.40	(.31)	.20
Δ interventie: conditie 'Docent-gestuurd'	.66	(.24)	<.01
<i>Implementatie</i>			
percentage gecontroleerde antwoorden ^a	.016	(.008)	.045
<i>Varianties</i>			
S ² klas	.31	(.13)	
S ² leerlingen: voortoets	12.71	(.74)	
S ² leerlingen: tussentoets	10.45	(.64)	
S ² leerlingen: natoets cloze	9.13	(.59)	
S ² leerlingen: natoets begripvragen	3.75	(.25)	

^a percentage: gecentreerd rond gemiddelde (88%)

Gemiddeld controleerden leerlingen 88% van hun antwoorden. Leerlingen die 100% van hun hun antwoorden controleren tijdens de lessenserie, scoren dus $.19 = (100 - 88) * .016$ hoger op de toetsen ná de interventie dan de gemiddelde leerlingen. Andersom geldt: leerlingen die 50% van hun antwoorden controleren tijdens de lessenserie, scoren ná de interventie $.61 = (88 - 50) * .016$ lager dan gemiddelde leerlingen, en $.80 = (100 - 50) * .016$ lager dan leerlingen die 100% controleren. Ook hier geldt dat dit effect relatief beperkt is: drie procent van de variantie tussen leerlingen kan ermee worden verklaard ($ES = .03$, $sd = .01$). In Bijlage 6 is de gemiddelde score van leerlingen op de vier toetsmomenten gevisualiseerd weergegeven in drie figuren.

Samenvattend. We kunnen dus aantonen dat leerlingen in de conditie 'Docent-gestuurd' leren in het algemeen van een wat hoger tekstbegrip blijkgeven ná de interventie op basis van zowel een cloze- als een begripvragentoets. Daarnaast geldt, onafhankelijk van conditie: naarmate leerlingen hun antwoorden vaker controleren, bouwen ze meer tekstbegrip op ná de interventie, dan dat ze vóór de lessenserie lieten zien; naarmate ze minder vaak hun antwoorden controleerden, bouwen ze minder tekstbegrip op ná de interventie.

3.4 Samenvatting resultaten per onderzoeksvraag

Vooropgesteld: blijkens de beschikbare data hebben docenten en leerlingen zich gehouden aan de instructie bij beide condities. De waardering voor de lessenserie door leerlingen en docenten was over het algemeen duidelijk positief.

De eerste onderzoeksvraag luidde: 'In hoeverre maakt de lessenserie vwo 4-leerlingen bewust van (de voorwaarden voor) diep begrip bij zakelijke teksten, en wijkt dit af per conditie?' Uit de evaluatie van leerlingen en docenten en vooral uit de resultaten op de *post-then-pre*-test kunnen we concluderen dat de lessenserie leerlingen hiervan inderdaad bewustmaakt. Deelnemers zelf geven ná de lessenserie aan dat ze vaker diep-leesgedrag zijn gaan inzetten dan vóór de lessenserie: naar eigen zeggen herlezen ze vaker tekstelementen, bouwen ze vaker (diep) begrip op en zetten ze vaker strategieën in. Dit geldt sterker voor de conditie 'Docent-gestuurd leren' dan voor de conditie 'Zelfstandig leren': leerlingen die de observerend-leerlessen klassikaal hebben doorgewerkt, ervaren dus een grotere verandering in hun leesgedrag dan leerlingen die deze lessen individueel hebben doorgewerkt.

De tweede onderzoeksvraag luidde: 'In hoeverre geven vwo 4-leerlingen die de lessenserie hebben gemaakt, blijk van meer diep begrip als ze daarna zelf zakelijke teksten lezen, en wijkt dit af per conditie?' Leerlingen in de conditie 'Docent-gestuurd leren' blijken aantoonbaar over het algemeen baat te hebben gehad van de interventie: zij blijken dan een wat hogere tekstbegripsscore te behalen dan dat ze dat vóór de lessenserie deden; voor de conditie 'Zelfstandig leren' kunnen we dit niet in het algemeen aantonen.

Kijken we naar de derde onderzoeksvraag: 'In hoeverre is de wijze van uitvoering van deze lessenserie van invloed op de effectiviteit ervan?' Er is duidelijk een effect aantoonbaar voor zowel de clozetest als de test met vragen, waarbij de wijze van uitvoering door de leerling wel een sterke invloed heeft. Voor beide condities geldt: naarmate leerlingen vaker hun gegeven antwoorden controleerden tijdens de lessenserie, leren ze meer erdoor, onafhankelijk van conditie. Leerlingen die meer dan gemiddeld hun antwoord controleren, hebben wél baat bij de lessenserie - terwijl leerlingen die minder dan gemiddeld hun antwoord controleren, er *niet* op vooruitgaan.

4 Discussie en conclusie

In deze studie onderzochten we onder 27 vwo 4-klassen, in twee condities, de effectiviteit van een lessenserie over diep begrip op a) het leesgedrag van leerlingen, zoals ervaren door henzelf, en op b) hun tekstbegrip, zoals gemeten in een viertal toetsen. Ontwerpprincipes bij de lessenserie vormden het diep lezen van zakelijke teksten, een procesgerichte benadering van tekstbegrip via observerend leren, flexibele strategie-instructie en een functionele inbedding van tekstbegrip.

In het algemeen geldt dat leerlingen na de lessenserie zelf een duidelijke verandering in hun eigen leesgedrag door de lessenserie rapporteren: ze herlezen naar eigen zeggen naderhand vaker tekstelementen, bouwen vaker (diep) begrip op, en zetten vaker strategieën in. Dit alles geldt echter aanmerkelijk sterker voor de conditie 'Docent-gestuurd leren' dan voor de conditie 'Zelfstandig leren'. Bovendien: als we niet verdisconteren hoe leerlingen de lessenserie hebben doorgewerkt, geven alleen leerlingen uit de conditie 'Docent-gestuurd leren' ná de lessenserie

gemiddeld blijkt van een wat hoger tekstbegrip dan vóór de lessenserie; in de conditie ‘Zelfstandig leren’ kunnen we dit effect niet aantonen. Niet geheel onverwacht blijkt de docent daarmee een cruciale rol te spelen in het creëren van een leereffect bij deze lessenserie (vgl. Hattie, 2012).

In de conditie waarin leerlingen zelfstandig, zonder begeleidende docent, de lessenserie doorwerken, wordt juist de wijze doorslaggevend waarop de leerling de lessen uitvoert. Voor beide condities geldt: leerlingen die consequent de tijd namen hun antwoord op gesloten vragen te controleren – leerlingen waarschijnlijk die over het algemeen wat meer gemotiveerd of geconcentreerd of gedisciplineerd te werk zijn gegaan –, blijken meer geleerd te hebben van de lessenserie dan de leerlingen die doorgaans niet de tijd hiervoor namen – dan leerlingen waarschijnlijk die over het algemeen wat minder gemotiveerd of geconcentreerd of gedisciplineerd te werk zijn gegaan. Deze bevinding kan aansluiten bij een bekende beperking van e-learning: in situaties waarin de intrinsieke motivatie van leerling om te leren niet vanzelfsprekend is, worden leerkenmerken (concentratie en motivatie, zelfdiscipline en *self-efficacy beliefs*, etc.) vaak doorslaggevend om te leren (Bouhnik & Marcus, 2006). Opvallend genoeg zien we hierin geen verschil optreden tussen condities: of de docent wel of niet het leerproces begeleidt, de wijze van uitvoering door een leerling lijkt in beide condities hetzelfde effect te hebben op de leerwinst.

4.1 Kanttekeningen

Vooropgesteld: deze interventiestudie is afgenomen in de eerste maanden van het schooljaar 2021-2022. Deze periode was door de aanhoudende Covid19-epidemie voor docenten en leerlingen vaak onoverzichtelijk en chaotisch: sommige docenten gaven in logboeken en interviews aan dat ze hun lessen als relatief rommelig ervoeren en dat ze meer dan in andere schooljaren leeroutine en -discipline bij hun vwo 4-leerlingen misten. De uitzonderlijke afnameomstandigheden leidden ook tot meer uitval dan gebruikelijk, op vrijwel alle onderdelen van het experiment.³ In hoeverre de omstandigheden daarmee een impact hebben gehad op de resultaten in deze studie, is een open vraag. Wel denken we de impact van de uitval op de resultaten te hebben kunnen minimaliseren door de gekozen statistische benaderingswijze en door de relatief omvangrijke steekproef.

Los van de afnameomstandigheden, zijn er de nodige voorbehouden bij deze studie te plaatsen. Zo hebben we in deze studie enkel via een vragenlijst gemeten in hoeverre leerlingen door de lessenserie een verandering in hun leesgedrag hebben ervaren. De vraag in hoeverre leerlingen hun leesgedrag ook in de praktijk hebben bijgesteld door de interventie, hebben we niet onderzocht via bijvoorbeeld oogbewegingsopnamen (vgl. bijv. Salmerón et al., 2020).

³ Een ander veelzeggend detail in dit opzicht: drie docenten van een achtste deelnemende school haakten enkele dagen voor afname van het experiment af, omdat ze door lesuitval en tijdsdruk meenden zich niet afdoende ervoor te kunnen inzetten.

Verder is het goed hier te signaleren dat we tal van leer- en leerlingkenmerken niet in het onderzoek verdisconteerd hebben, kenmerken die in andere studies wel een interessant perspectief op de resultaten boden, zoals het onderscheid tussen vaardige en minder vaardige lezers (bijv. Van Silfhout, 2014), de *self-efficacy beliefs* van leerlingen (bijv. Elving-Heida, 2019) en hun metacognitieve vaardigheden en kennis van tekstconventies, -kenmerken en -genres (bijv. Bouwer & Koster, 2016). De vraag in hoeverre sommige leerlingen meer van de interventie hebben geprofiteerd dan andere leerlingen, blijft in deze studie onderbelicht. Verder geldt dit onderzoek slechts vwo 4-leerlingen; in hoeverre de lessenserie ook effectief is voor andere doelgroepen (havo- en mavo-leerlingen bijvoorbeeld), hebben we niet onderzocht.

Bovendien moet hier worden opgemerkt dat we niet nader hebben onderzocht welke componenten in de lessenserie de effectiviteit ervan kunnen verklaren. In hoeverre onze studie direct de effectiviteit van observerend leren onderbouwt, in het bijzonder voor de leesles (vgl. Couzijn, 1995), valt in strikte zin niet te zeggen. We hebben immers niet onderzocht in hoeverre specifiek de component 'observerend leren' in de lessenserie dit positieve resultaat heeft veroorzaakt; mogelijk kunnen andere componenten in de lessenserie (bijv. de functionele inbedding van tekstbegrip of het weghalen van de tekst na bestudering) evenzeer de werkzaamheid van onze lessenserie verklaren.

4.2 Wetenschappelijke implicaties

Met de experimentele lessenserie in onze studie ontwikkelden we een grotendeels procesgerichte benadering van de les tekstbegrip, waarbij we aansloten bij observerend leren (bijv. Bandura & Walters, 1977; Bandura, 1986), bij recente inzichten rond de intentionaliteit van tekstbegrip (bijv. Britt et al., 2018; McCrudden et al., 2011), bij een brede conceptualisatie van evaluatie en beoordeling van teksten (OECD, 2019), en bij inzichten rondom (de kenmerken van) diep begrip (Wolf, 2018).

In een aantal opzichten heeft deze studie wetenschappelijke implicaties. Allereerst kan ze – zoals gezegd – als een mogelijke onderbouwing gelden voor de effectiviteit van observerend leren voor het leesonderwijs, waarnaar nog relatief beperkt onderzoek bestaat, nationaal (zie hoofdstuk 5; Couzijn, 1995; Keehnen et al., 2015) en internationaal (Salmerón & Llorens, 2019; Salmerón et al., 2020). Ten tweede: in onze experimentele lessenserie worden recente theoretische inzichten geïntegreerd over intentioneel tekstbegrip, over het belang van evaluatie en beoordeling en over diep begrip. Naar de mogelijkheden om deze inzichten afzonderlijk, in specifieke interventies in het klaslokaal te vertalen, bestaat al het nodige onderzoek: eerder werd bijvoorbeeld aangetoond dat leerlingen via observerend leren effectief leren reflecteren en evalueren aangaande de kwaliteit van bronnen (bijv. Braasch et al., 2013; Wiley et al., 2013). Onze studie verkent hoe leerlingen vanuit een intentioneel kader (kenmerken van) diep begrip kan worden onderwezen: daarmee is het een van de eerste interventiestudies die deze moderne inzichten in tekstbegrip *in samenhang* tot een effectieve, sterk procesgerichte lessenserie voor scholier-

ren heeft geïntegreerd. Ten derde bevestigt deze studie andermaal de belangrijke rol van interactie bij e-learning (vgl. Bouhnik & Marcus, 2006; Martínez et al., 2007; Udo et al., 2011): zonder de begeleidende rol van de docent en de interactie met klasgenoten lijken vooral leerlingkenmerken te bepalen in hoeverre leerlingen tot leren komen bij onze lessenserie.

4.3 *Implicaties voor het onderwijs*

Daarnaast heeft deze studie interessante implicaties voor het leesonderwijs binnen het schoolvak Nederlands. Er loopt momenteel een discussie in Nederland of het onderwijs in tekstbegrip in het voortgezet onderwijs niet aan herziening toe is (bijv. SLO, 2021b). In deze studie wordt tekstbegripsonderwijs duidelijk anders benaderd dan op veel Nederlandse voortgezet-onderwijsscholen in het schoolvak Nederlands momenteel de standaard lijkt: de traditionele tekst-met-vragen-aanpak (vgl. Frank et al., 2019). Zowel deelnemende docenten als leerlingen gaven aan deze afwijkende benadering te herkennen én doorgaans ook te waarderen: het leeuwendeel van docenten vond de benadering interessant en leerzaam (“erg tof”, “een eye-opener”). Deze afwijkende benadering lijkt zich ook te hebben vertaald in een andere leesaanpak: leerlingen geven na de lessenserie aan dat ze vaker diepleesgedrag zijn gaan inzetten dan vóór de lessenserie; bovendien lieten leerlingen na de lessenserie meer tekstbegrip zien, al is dit effect relatief klein.

In veel opzichten contrasteert de lessenserie in deze studie met de reguliere tekst-met-vragen-didactiek. Deze traditionele didactiek wordt weliswaar allang scherp bekritiseerd (Durkin, 1978), maar heeft nog altijd breed ingang in het voortgezet onderwijs (bijv. Bogaerds-Hazenbergh et al., 2022b; Hoogeveen & Bonset, 1998; Westhoff, 2012; etc.). In de reguliere benadering staat vooral het product centraal, namelijk de leerlingantwoorden op tekstafhankelijke vragen, en blijft het leesproces van leerlingen doorgaans onbesproken; in de observerend-leren-lesserie in onze lessenserie richten leerlingen zich daarentegen geheel op het leesproces, vanuit de vraag in hoeverre het leesgedrag van leeftijdsgenoten in filmpjes tot diep begrip leidt. In de reguliere aanpak worden vaak strategieën context- en doeleloos aangeboden, als generieke ‘manieren van lezen’, en worden evidencebased leesstrategieën enkel indirect aangeboden (Rooijackers et al., 2021b); in onze lessenserie staat daarentegen het doel centraal, namelijk het opbouwen van diep begrip voor een schoolse taak, en wordt een flexibele, doelmatige inzet van evidencebased strategieën voorgestaan. In de reguliere aanpak ontbreekt vrijwel steeds een functionele context en worden leerlingen vrijwel steeds sets tekstafhankelijke vragen aangeboden, die elkaar in een schijnbaar willekeurige volgorde afwisselen; in onze lessenserie zijn daarentegen alle leestaken ingebed binnen een functioneel schooldoel, wisselen de leestaken sterk in vorm, en in de werklessen sluiten de vervolgtaken logisch aan op dit doel.

Met deze afwijkende benadering van de les tekstbegrip kan de lessenserie aansluiten bij een aantal actuele thema’s binnen het schoolvak Nederlands. Allereerst: andermaal toont deze studie dat een meer procesgerichte benadering via obser-

verend leren bruikbaar én effectief kan zijn in het domein leesvaardigheid in het schoolvak Nederlands (vgl. hoofdstuk 5; Couzijn, 1995; Keehnen et al., 2015). De werkwijze in de leesles Nederlands lijkt vaak bij uitstek productgericht; de vraag is of het Nederlandse leesonderwijs niet sterker op de metacognitieve en zelfregulerende vaardigheden van leerlingen in zou moeten spelen door het leesgedrag van leerlingen zelf centraal te stellen.

De lessenserie sluit daarnaast aan bij de roep om meer diep lezen in het Nederlandse onderwijs en verdisconteert evalueren en reflecteren in tekstbegripsonderwijs, dimensies van tekstbegrip waaraan het Nederlandse leesonderwijs nog onvoldoende aandacht lijkt te schenken (Gubbels, 2020; OECD, 2019; Onderwijsraad & Raad voor Cultuur, 2019). De lessenserie kan voorts aansluiten bij de roep om meer 'bewuste geletterdheid' (Meesterschapsteam, 2019) en meer taalbewustzijn bij Nederlandse leerlingen in het voortgezet onderwijs (curriculum.nu, 2019): leerlingen blijken na de lessenserie immers een duidelijke bewustzijnsverandering in hun leesgedrag te ervaren. Ten slotte: doordat de lessenserie tekstbegrip steeds inkadert binnen schoolse taken en daarmee tekstbegrip inbedt binnen andere taalvaardigheden (schrijfvaardigheid, debatteren, etc.), sluit ze aan bij de roep om ont-kaveling binnen het schoolvak Nederlands (Van Gelderen & Van Schooten, 2011; SLO, 2021b).

4.4 Aandachtspunten

Op een drietal punten biedt deze studie stof ter overweging. Allereerst: in eerdere studies werd al getoond hoezeer Nederlandse leerlingen grote taakvertrouwdheid en taakafhankelijkheid bij de traditionele tekst-met-vragen hebben opgebouwd (zie hoofdstuk 3; Bogaerds-Hazenbergh et al., 2022b); ook in deze studie lijkt vertrouwdheid met deze standaardaanpak haar sporen na te laten. In de werklessen van de lessenserie werd bij leestaken erop gehamerd dat leerlingen vooraf een tekst diepgaand moesten bestuderen en deze daarna, bij de verwerkingstaak, niet meer moesten raadplegen, zodat ze tijdens de verwerkingstaak konden controleren of ze tijdens bestuderen voldoende diep begrip hadden opgebouwd. Daarmee wilden we bereiken dat leerlingen vooraf zelfstandig een tekst leerden bestuderen – iets waartoe de reguliere tekst-met-vragen-aanpak niet lijkt aan te zetten (zie vooral hoofdstuk 4). Echter, een kleine 40% van de deelnemende leerlingen in het experiment gaf aan dat ze deze instructie hadden genegeerd. Daarmee vielen ze terug op het hun vertrouwde scenario 'lees de taak, zoek daarna het antwoord in de tekst'. Ook in de evaluatie door docenten en leerlingen klonken er met enige regelmaat minder positieve geluiden die aan de standaardaanpak zijn te verbinden en onwennigheid met de afwijkende benadering suggereerden. (Vergelijk het leerlingantwoord: "Dit was niet interessant voor mij, [het] helpt me niet met de gewone les, we hebben voor mijn gevoel de helft van deze lessen naar mensen die lezen gekeken.")

Tweede overweging: deze studie lijkt weliswaar de effectiviteit van observerend leren aan te tonen – al hebben we dit niet nader onderzocht: zie onze opmerking in § 4.1 –, een open vraag is wel in hoeverre erin de potentie ervan afdoende is benut.

Om voldoende diepgang in hun reflectie te kunnen opbouwen, waren leerlingen en docenten geïnstrueerd elk modelimpje vooral twee keer te bekijken; deze instructie was gebaseerd op ervaringen uit eerdere try-outs. In het experiment lijkt ze echter niet breed te zijn nageleefd: docenten en leerlingen hielden zich in de geobserveerde lessen er in elk geval doorgaans niet aan. Omdat we de leerlingantwoorden in de observerend-leren-lessen niet uitgebreid kwalitatief hebben onderzocht, valt niet te zeggen in hoeverre deze en andere afwijkingen van de instructie de evaluatie en reflectie op de modellen hebben beïnvloed, en daarmee ook het leereffect van de lessenserie. Kwalitatief vervolgonderzoek naar de voorwaarden voor effectiviteit van observerend leren lijkt hoe dan ook in het algemeen raadzaam (vgl. Van der Loo, 2022).

Derde overweging, samenhangend met het vorige punt: de vraag is in hoeverre docenten niet nader geschoold en getraind zouden moeten worden in een meer procesgerichte benadering van tekstbegrip. Mede doordat docenten vooraf geen uitgebreide inhoudelijke of didactische training hadden ontvangen door de onderzoekers, ontstonden er waarschijnlijk grote verschillen in de aanpak van docenten. Sommige docenten in de conditie ‘Docent-gestuurd leren’ gaven in hun logboeken aan dat ze levendige, inhoudelijke klassengesprekken hielden, waar andere docenten aangaven dit proces moeizamer te vinden verlopen. Een enkele docent uit deze conditie bleek in het geheel niet overtuigd van het nut van deze werkwijze en achtte de lessenserie “geen goede voorbereiding op het Centraal Examen Nederlands” (wat inderdaad geen doel ervan was). Een zekere mate van training of scholing lijkt al met al wenselijk, al heeft de gevarieerde aanpak van de docent de algemene effectiviteit van de lessenserie in de conditie ‘Docent-gestuurd leren’ niet in de weg gestaan.

4.5 Conclusie

In een tijd waarin leerlingen hun informatie vooral via internet en sociale media tot zich nemen, lijkt er alle reden om leerlingen in het voortgezet onderwijs nadrukkelijk te confronteren met een afwijkende diepgaande, aan close-reading verwante leesaanpak en de voordelen die zo’n aanpak voor het leerproces kan hebben (vgl. Onderwijsraad & Raad voor Cultuur, 2019; zie ook Lapp et al., 2018). Dat geldt niet in het minst voor Nederlandse vwo-leerlingen die in studie en werk vaak te maken zullen krijgen met diep-lezen en de intensieve, herhaalde leeswijze die dit vergt. Huidige Nederlandse universitaire studenten zouden nu over het algemeen te weinig kritisch kunnen lezen (Warps et al., 2021). Met deze uitgebreid geconstrueerde en vooraf uitgeprobeerde lessenserie kan vwo-leerlingen een ‘introductie-cursus’ op het concept diep lezen worden aangeboden.

Enkele van de ontwerpprincipes achter deze lessenserie zijn eenvoudig door docenten in het klaslokaal te vertalen. Te denken valt aan: a) het functioneel inkaderen van leestaken, bijvoorbeeld door leestaken te integreren in presentatie-, discussie-, en schrijftaken, b) het procesgericht benaderen van tekstbegrip, bijvoorbeeld door leerlingen onderling hun leesproces te laten bespreken en evalueren, c) het afwis-

selend aanbieden van soorten leestaken, zoals sorteertaken, samenvattaken, argumentatietaken, etc., en d) het benadrukken van het belang van de zelfstandige tekstbestudering door een leerling, bijvoorbeeld door de bestudeerde tekst in te nemen bij het uitreiken van de verwerkingstaak.

We hopen te gelegener tijd deze lessenserie openbaar toegankelijk te maken voor gebruik in de les; mogelijk kan ze ook door educatieve uitgevers verder worden uitgewerkt. Het zelfstandig construeren van een lessenserie als deze is immers voor docenten nogal complex: in onze lessenserie spelen oogbewegingsfilmpjes bijvoorbeeld een belangrijke rol; Nederlands middelbare scholen beschikken momenteel niet over een eye-track-computer waarmee zulke EMME's kunnen worden geconstrueerd. In een aansluitende, alternatieve lessenserie zouden uitgevers dan aspecten van tekstbegrip kunnen uitdiepen die vooralsnog redelijk beperkt in de lessenserie zijn uitgewerkt, zoals meervoudig-tekstbegrip (Britt & Rouet, 2012) en de bruikbaarheid en betrouwbaarheid van informatie (Kirschner, 2017).

Kortom: we denken met deze lessenserie een alternatieve, effectieve en zinvolle blik op tekstbegripsonderwijs te hebben laten zien, met meer oog voor het leesgedrag van leerlingen en de effectiviteit ervan, voor reflectie en evaluatie, en voor functioneel, betekenisvol lezen. De lessenserie in onze studie kan zo een inspiratiebron vormen voor een effectieve, meer eigentijdse vakinhoudelijke en -didactische benadering van het domein leesvaardigheid.

Met de resultaten ervan denken we bovenal andermaal de cruciale toegevoegde waarde van de docent in de les te hebben laten zien (vgl. Hattie, 2012). In de twee lockdowns tijdens de Covid19-epidemie werd deze meerwaarde in het voortgezet onderwijs weleens ter discussie gesteld doordat het thuiswerken van leerlingen tijdelijk de norm was (Smeets, 2021). Leergangen Nederlands voor het voortgezet onderwijs zetten in hun leesmodules inmiddels al enige tijd deels in op e-learning (vgl. bijv. Frank et al., 2019). Het is echter de vraag of dergelijke vormen van e-learning bij middelbarescholieren over het algemeen zinvol zijn. Deze lessenserie laat juist zien dat in leersituaties *zonder* een docent vooral leerkenmerken van leerlingen bepalend worden of ze tot leren komen. In een samenleving zoals de Nederlandse, die sterk waarde hecht aan kansgelijkheid (bijv. Onderwijsraad, 2019), kan beter niet overmatig worden vertrouwd op onderwijs *z*onder de sturende rol van de docent.

HOOFDSTUK 7

Discussie

1 Doel en opzet van het onderzoek

In de inleiding van dit boek schetsten we een aantal aandachtspunten rond tekstbegrip en het schoolvak Nederlands. De inhoud en didactiek in huidige leergangen Nederlands voor vwo sluiten in belangrijke opzichten nog aan bij inzichten uit het schoolvak Nederlands in de jaren zeventig en tachtig van de vorige eeuw. Daarnaast is er relatief beperkt onderzoek naar de vaardigere lezers in het Nederlandse voortgezet onderwijs, vwo-leerlingen. Nederlandse leerlingen lijken verder internationaal gezien minder sterk in evalueren en reflecteren op teksten, waardoor er binnen het schoolvak Nederlands een roep weerklinkt aan deze diepere begripsdimensies in de leeslessen meer aandacht te schenken (bijv. Onderwijsraad & Raad voor Cultuur, 2019). Bovendien is er een inhoudelijk herzieningstraject van start gegaan ten aanzien van het schoolvak Nederlands in de bovenbouw, waarin ook het leesonderwijs een prominent aandachtspunt vormt (Rijksoverheid, 2022; SLO, 2021b). We hebben daarom vastgesteld dat er alle reden is om stil te staan bij het tekstbegrip van vaardige lezers, vwo-leerlingen, binnen de huidige reguliere context van het schoolvak Nederlands.

In de inleiding hebben we ons twee onderzoeksdoelen gesteld: enerzijds wilden we nauwkeuriger het tekstbegrip van vwo-leerlingen in de bovenbouw beschrijven, door hun vooraf-lees- en antwoordgedrag te beschrijven bij de tekst met vragen, de courante leestaak in de les zakelijk lezen in het schoolvak Nederlands; anderzijds wilden we een alternatieve effectieve invulling ontwikkelen en beproeven van de lessen tekstbegrip voor leerlingen in de vwo bovenbouw.

Daarbij vertrokken we vanuit zeven uitgangspunten: 1) We wilden ons onderzoek richten op diep begrip. 2) Daarbij vormden vaardige lezers, vwo-leerlingen, de onderzoeksgroep. 3) Het indirecte, nogal ongrijpbare verschijnsel tekstbegrip wilden we zo tastbaar mogelijk maken door het te visualiseren: we gebruiken oogbewegingsonderzoek als beschrijvend onderzoeksinstrument en oogbewegingsfilmpjes als didactisch instrument, zodat tekstbegrip als het ware wordt veraanschouwelijkt voor onderzoekers, lezers zelf én voor leerlingen. 4) Het leesproces en het antwoordproces bij leestaken wilden we daarbij nadrukkelijk van elkaar onderscheiden. 5) Het strategisch leesgedrag van lezers zou daarbij steeds doelmatig worden benaderd: de vraag staat centraal in hoeverre lezers erin slagen met hun leesmiddelen het leesdoel te realiseren. 6) En bij dit alles streefden we naar een optimale ecologische validiteit en 7) – meer impliciet – naar een optimale generaliseerbaarheid.

Vanuit de twee onderzoeksdoelen heeft dit boek een tweeledige opbouw. De eerste helft, hoofdstuk 2 tot en met 4, is beschrijvend en deels evaluerend, waarbij we vooral een *mixed method*-benadering hebben gebruikt. In hoofdstuk 2 en 3 hebben we het vooraf-lees- en antwoordgedrag van vwo-leerlingen bij teksten met vragen beschreven en geanalyseerd, en in hoofdstuk 4 hebben we dit vooraf-lees- en antwoordgedrag gemanipuleerd. Zo zijn we op een aantal pijnpunten gestuit in de huidige dominante didactiek in tekstbegrip binnen het schoolvak Nederlands.

De tweede helft van dit boek, hoofdstuk 5 en 6, heeft betrekking op ontwerp- en interventieonderzoek, waarbij we met name hebben aangesloten bij de leerprincipes van observerend leren (Bandura & Walter, 1977; Bandura, 1986) en bij recente theoretische inzichten rondom tekstbegrip (bijv. Britt et al., 2018). Op grond hiervan hebben we in deze hoofdstukken alternatieve ontwerpprincipes opgesteld voor de lessen tekstbegrip en hebben we een alternatieve benadering ontworpen en gevalideerd voor de lessen tekstbegrip in met name de vwo bovenbouw – een benadering die in grote lijnen overigens evenzeer voor andere afdelingen in het voortgezet onderwijs bruikbaar lijkt.

De resultaten uit de vijf studies in hoofdstuk 2 tot en met 6 zullen in dit slothoofdstuk uitgebreid en systematisch worden hernomen. Dit slothoofdstuk bestaat uit vier delen. Allereerst beantwoorden we in § 2 de drie centrale onderzoeksvragen die we ons in de inleiding hebben gesteld. Daarna bespreken we in § 3 de wetenschappelijke implicaties van hoofdstuk 2 tot en met 6 én doen we suggesties voor vervolgonderzoek. Vervolgens gaan we in § 4 uitgebreid in op de implicaties van ons onderzoek voor het onderwijs, en dan met name voor de lessen tekstbegrip voor vwo(-bovenbouw)-leerlingen. Ten slotte eindigen we in § 5 met een persoonlijke algemene slotopmerking.

2 De centrale onderzoeksvragen beantwoord

In de inleiding hebben we onszelf drie centrale onderzoeksvragen gesteld die we door middel van vijf deelstudies hebben beantwoord:

- a) Hoe bouwen vwo-leerlingen tekstbegrip op en welke ontwikkeling maken ze hierin door, als we uitgaan van de reguliere praktijk in de lessen tekstbegrip in het schoolvak Nederlands?
- b) In hoeverre vormt de tekst met vragen een effectieve taak om tekstbegrip te oefenen in de lessen tekstbegrip in het schoolvak Nederlands?
- c) Wat is een effectieve, procesgerichte inhoud en didactiek voor de lessen tekstbegrip in het schoolvak Nederlands?

In het onderstaande zullen we deze drie onderzoeksvragen beantwoorden, op grond van de resultaten uit de studies in hoofdstuk 2 tot en met 6.

2.1 Eerste onderzoeksvraag

We starten met de eerste centrale onderzoeksvraag:

- a) Hoe bouwen vwo-leerlingen tekstbegrip op en welke ontwikkeling maken ze hierin door, als we uitgaan van de reguliere praktijk in de lessen tekstbegrip in het schoolvak Nederlands?

In het algemeen schragen de resultaten uit hoofdstuk 2 en 3 de wettelijke beschrijvingen van het huidige beheersingsniveau voor het lezen van zakelijke teksten door vwo-leerlingen, zoals vastgelegd in het Referentiekader Taal – dat overigens recent geëvalueerd is (zie Van den Broek et al., 2022). De veronderstelling in het Referentiekader Taal dat leerlingen vanaf niveau 2F hoofdzaken van bijzaken kunnen scheiden (vgl. Expertgroep Doorlopende Leerlijnen Taal, 2008, 2009), wordt in hoofdstuk 2 bevestigd: bij de reguliere tekst met vragen bleken de vwo 4-leerlingen over het algemeen kernzinnen te onderscheiden van niet-kernzinnen. Ook in hoofdstuk 3 zagen we deze veronderstelling bevestigd.

In hoofdstuk 3 zagen we daarnaast cross-sectionele verschillen tussen lagere en hogere jaarlagen die een duidelijke ontwikkeling in leesvaardigheid suggereren conform de beschrijving in het Referentiekader Taal. Vwo 4- en vwo 6-leerlingen lezen blijkens dit hoofdstuk gemiddeld beduidend sneller dan vwo 2-leerlingen en ze herlezen vaker tekstdelen: ze lezen ogenschijnlijk enigszins strategischer dan vwo 2-leerlingen. Tussen vwo 4- en vwo 6-leerlingen kunnen we dergelijke verschillen echter opvallend genoeg niet aantonen. Deze bevindingen lijken te sporen met de beschrijving in het Referentiekader Taal, waar de inhoudelijke stap van basisniveau 2F (min of meer: vwo 2) naar het meer verdiepende niveau 3F (min of meer: vwo 4) groter lijkt dan de stap van niveau 3F naar 4F (vwo 6): er ontbreken bijvoorbeeld nogal eens differentiaties tussen niveau 3F en 4F, die wel te vinden zijn tussen niveau 2F en 3F (vgl. Van den Broek et al., 2022), en de examensyllabi die het examenprogramma voor tekstbegrip beschrijven voor 3F (havo) en 4F (vwo), onderscheidden zich inhoudelijk nauwelijks van elkaar (CvTE, 2022a, 2022b).

Daarnaast viel het in hoofdstuk 3 op dat de drie onderzochte jaarlagen zich niet zozeer in hun vooraf-leesgedrag onderscheiden: zowel vwo 2-, als vwo 4-, als vwo 6-leerlingen lezen over het algemeen nogal lineair, met relatief weinig herlezingen en met een ogenschijnlijk weinig actieve inzet van (vooraf-)leesstrategieën. Een van de kerndoelen Nederlands luidt dat leerlingen aan het einde van de onderbouw van het voortgezet onderwijs in staat zouden moeten zijn strategieën te gebruiken en te beoordelen (SLO, 2016a; Expertgroep Doorlopende Leerlijnen Taal, 2008); de vwo-leerlingen in onze steekproeven lijken echter over het algemeen slechts opvallend beperkt strategieën in te zetten bij het vooraf-lezen van teksten, ongeacht jaarlaag en ongeacht de complexiteit van de teksten, die in niveau varieerden van niveau 2F tot 4F. Daarmee is niet gezegd dat vwo-leerlingen deze strategieën niet in kunnen zetten; eerder lijken ze door de leestaken in onze experimenten niet te zijn aangezet tot dergelijk intensief strategisch leesgedrag.

Leerlingen uit de drie onderzochte jaarlagen onderscheidden zich in hoofdstuk 3 daarentegen wel zeer duidelijk in hun antwoordproces; dat is in vergelijking met het vooraf-leesproces doorgaans (zeer) omstandig en bestaat meestal uit vele handelingscomponenten. Vooral het onderscheid tussen vwo 2 en vwo 6 is hier markant: vwo 6-leerlingen lezen allereerst de vraag korter dan vwo 2-leerlingen – omdat ze deze waarschijnlijk beter kunnen interpreteren; vwo 6-leerlingen hebben na het lezen van de vraag vaker al een idee van het antwoord – omdat ze waarschijnlijk een beter situatiemodel van de tekst hebben opgebouwd; vwo 6-leerlingen weten bij de vraagbeantwoording de antwoord-relevante tekstdelen beter te vinden en besteden hun tijd minder én gericht aan antwoord-irrelevante tekstdelen dan vwo 2-leerlingen – omdat ze vooraf waarschijnlijk een betere tekstbasis van de tekst hebben opgebouwd; en, *last but not least*, vwo 6-leerlingen hebben uiteindelijk vaker een goed antwoord – omdat ze vooraf waarschijnlijk een betere mentale representatie van de tekstinhoud hebben opgebouwd en vervolgens een verfijnder en adequater antwoordprocédé doorlopen. Vwo 4-leerlingen schuren in hun gedrag hierbij vaker tegen vwo 6-leerlingen aan dan tegen vwo 2-leerlingen – waarmee andermaal het beeld wordt bevestigd dat de stap van 2F (vwo 2) naar 3F (vwo 4) groter is dan die van 3F naar 4F (vwo 6).

Selectie vormt voor deze verschillen tussen jaarlagen overigens zeker een verklaring: door afstroom en doublure is de vwo 6-populatie immers homogener dan de vwo 2-populatie en (in mindere mate) de vwo 4-populatie. Desalniettemin suggereert deze cross-sectionele beschrijving van lees- en antwoordgedrag in veel opzichten een duidelijke ontwikkeling in leesvaardigheid van vwo-leerlingen over jaarlagen heen, die in grote lijnen correspondeert met de beschrijving in het Referentiekader Taal. Voor de goede orde: deze ontwikkeling blijkt dan vooral uit de beantwoording van de vragen, niet zozeer uit het vooraf-leesproces. Pas tijdens het antwoordproces zien we dat het vrij lineaire, schijnbaar nogal oppervlakkige vooraf-leesproces bij de drie jaarlagen tot verschillende tekstbases en situatiemodellen heeft geleid.

2.2 Tweede onderzoeksvraag

Beantwoorden we dan de tweede centrale onderzoeksvraag:

- b) In hoeverre vormt de tekst met vragen een effectieve taak om tekstbegrip te oefenen in de lessen tekstbegrip in het schoolvak Nederlands?

De reguliere leestaak 'tekst met vragen' drong zich in hoofdstuk 2 en 3 nadrukkelijk op de voorgrond. We konden op grond van de eye-track-data in deze hoofdstukken hooguit een zwak verband aantonen tussen enerzijds het vooraf-lezen van de tekst en anderzijds het beantwoorden van de vraag en het uiteindelijke antwoord. In hoofdstuk 2 bleken vwo 4-leerlingen tijdens vooraf-lezen weliswaar onderscheid te maken tussen kernzinnen en niet-kernzinnen, maar een verband hierbij met het antwoordgedrag én met het antwoord konden we in het geheel niet aantonen. Deze bevinding replicateerden we grotendeels in hoofdstuk 3: hier konden we hooguit een zwak verband aantonen tussen (niet-)kernzinlezen en antwoord(gedrag).

Leerlingen lijken vooral vooraf te lezen omdat ze daartoe geïnstrueerd worden en omdat ze daardoor bij de vraagbeantwoording de vraag beter kunnen begrijpen en beter kunnen lokaliseren waar het antwoord in de tekst moet worden gezocht. Weliswaar zagen we in hoofdstuk 3 in de hardop-denk-data dat vwo 4- en vooral vwo 6-leerlingen na vraaglezing vaker dan vwo 2-leerlingen wisten waar ze het antwoord in de tekst moesten zoeken, maar slechts relatief zelden hadden leerlingen naar eigen zeggen direct na vraaglezing een (idee van een) antwoord in gedachten: leerlingen die vooraf langer stilstonden bij de belangrijkste zinnen óf korter bij de bijzaken in de tekst, vertoonden daarna nauwelijks of niet adequater antwoordgedrag en hadden ook niet aantoonbaar vaker een goed antwoord op de vragen, in vergelijking met leerlingen die vooraf korter stilstonden bij de belangrijkste zinnen of juist langer stilstonden bij de bijzaken.

Bovendien vonden we in hoofdstuk 3 sporen van sterk taakafhankelijk en taakvertrouwd gedrag: een kleine groep leerlingen gaf (ongevraagd!) aan zich bij de vraagbeantwoording geheel te laten leiden door conventies rondom de tekst met vragen (“De eerste vraag gaat meestal over alinea 1”, “De eerste vraag ging over alinea 1, dus de tweede vraag zal wel over alinea 2 of 3 gaan”, etcetera). Daarmee lijken de resultaten uit hoofdstukken 2 en 3 de vaak geuite kritiek op deze taak te onderbouwen dat leerlingen er vooral ‘trucjes’ mee leren en vooral leren hoe ze begripsvragen bij een tekst slim kunnen beantwoorden (bijv. Hoogeveen & Bonset, 1998; Westhoff, 2012).

Hoofdstuk 4 bevestigde de conclusie dat de antwoordfase bij deze taak cruciaal is: het draait vooral om het goed raadplegen van de tekst bij het maken van de vragen, en niet om een intensieve, nauwkeurige vooraf-bestudering van de tekst. In de studie in hoofdstuk 4 hebben we de intensiteit van het vooraf-lezen bij twee teksten met vragen gemanipuleerd, in twee experimenten. In het eerste experiment (‘Raadpleegbaarheid’) vroegen we ons af welke impact de raadpleegbaarheid van de tekst tijdens de antwoordfase heeft op de antwoordscore bij begripsvragen: als ze vooraf een tekst hebben bestudeerd, in hoeverre behalen de leerlingen die deze tekst bij de vragen kunnen raadplegen, een hogere score op de begripsvragen, dan de leerlingen die deze tekst bij de vragen niet langer kunnen raadplegen? In het tweede experiment (‘Intensiteit’) vroegen we ons af welke impact de intensiteit van vooraf-lezen heeft op de antwoordscore bij begripsvragen: in hoeverre behalen leerlingen die de tekst vooraf intensief hebben bestudeerd, een hogere score op de begripsvragen dan leerlingen die de tekst vooraf niet of minder hebben bestudeerd? Met deze experimenten wilden we dus achterhalen of vooral het vooraf-bestuderen of juist vooral de tekstraadpleging tijdens de vragen de antwoordscore bepaalt. Om dat te kunnen vaststellen, maakten we in beide experimenten gebruik van twee teksten met begripsvragen op situatiemodelniveau, twee toetsen die eerder als onderdelen van de Centraal Examens Nederlands havo/vwo hebben gefungeerd en min of meer als ‘exemplarische’ tekstbegripstoetsen kunnen gelden voor het Nederlandse tekstbegripsonderwijs (zie Canton et al., 2013).

In het eerste experiment (‘Raadpleegbaarheid’) zagen we dat leerlingen die tijdens de vraagbeantwoording niet langer de tekst konden raadplegen, beduidend

lagere scores haalden dan leerlingen die wel de tekst konden raadplegen: de tekst-raadpleegbaarheid heeft dus – niet onverwacht – een groot effect op de antwoordscore. In het tweede experiment ('Intensiteit') ontvingen drie groepen leerlingen wisselende instructies om de tekst al dan niet intensief te bestuderen, waaraan ze zich blijken onze controle-instrumenten hebben geconformeerd. Desalniettemin konden we in dit experiment – wél onverwacht – géén verschil in antwoordscore aantonen tussen deze drie groepen leerlingen.

Een uitgebreide vooraf-tekstbestudering draagt bij deze taak dus niet (of nauwelijks) bij aan een hogere antwoordscore bij de vragen; als de tekst bij de vragen raadpleegbaar blijft, dan bepaalt de tekstraadpleging vooral de antwoordscore. Daarmee ligt het voor de hand dat vaardige lezers zoals vwo-leerlingen vooraf een tekst enkel oppervlakkig zullen doorwerken. Ook vanuit recente theoretische inzichten (met name het RESOLV-model: Britt et al., 2018) verklaarden we in dit hoofdstuk de onverwachte (want niet aantoonbare) resultaten in dit tweede experiment: lezers zullen de tekst vooraf relatief oppervlakkig lezen om vragen adequaat te kunnen interpreteren en gemakkelijk de antwoord-relevante tekstdelen te kunnen vinden; de instructie een tekst *intensief* te bestuderen zullen ze grotendeels negeren, omdat pas bij de vragen het eigenlijke doel van de opdracht in beeld komt. Vooral de activiteiten in de antwoordfase bepalen dan in hoeverre leerlingen een goed antwoord op de vragen geven. In hoofdstuk 2 en 3 zagen we dit pragmatische vooraf-lees- en antwoordgedrag al bij de deelnemende vwo-leerlingen. Belangrijke voorwaarde hierbij is wel dat de tekst raadpleegbaar blijft. Blijft de tekst niet raadpleegbaar bij de vragen, zoals in het eerste experiment in hoofdstuk 4, dan zullen lezers de tekst juist wél intensief bestuderen, willen ze relatief complexe begripsvragen correct kunnen beantwoorden, zo beredeneerden we vanuit het model van Britt et al. (2018).

Onze conclusie luidde in hoofdstuk 4 daarom: wie leerlingen zelfstandig een tekst wil leren lezen, dient terughoudend met deze taak om te gaan; leerlingen leren ermee vooral goed begripsvragen bij een tekst te beantwoorden, maar zelfstandig lezen stimuleer je er waarschijnlijk niet mee. Op basis van ons onderzoek in hoofdstuk 2, 3 en 4 lijkt de huidige dominantie van deze taak in de lessen tekstbegrip niet erg voor de hand te liggen.

Hoewel deze conclusies in strikte zin slechts vwo-leerlingen gelden, lijkt er geen reden om ze niet voor het leesonderwijs in het algemeen te laten gelden: als vaardige lezers zoals vwo-leerlingen sterk pragmatisch lees- en antwoordgedrag vertonen, lijkt er geen reden aan te nemen dat relatief minder vaardige lezers (havo- en vmbo-leerlingen) niet eveneens zulk pragmatisch lees- en antwoordgedrag zullen vertonen. Hoogstwaarschijnlijk geldt de redenering – op grond van de gemiddeld lagere motivatie, concentratie en vaardigheid van minder vaardige lezers – bij deze lezers zelfs eerder *a fortiori*.

2.3 Derde onderzoeksvraag

Laten we ons dan richten op de derde centrale onderzoeksvraag in dit boek:

- c) Wat is een effectieve, procesgerichte inhoud en didactiek voor de lessen tekstbegrip in het schoolvak Nederlands?

In hoofdstuk 5 en 6 beschreven we een aantal uitgangspunten voor een alternatieve inhoud en didactiek voor de lessen tekstbegrip voor vwo-leerlingen. Vooropgesteld: we vertrokken in deze twee hoofdstukken vanuit het begrip diep lezen (Wolf, 2018): het geconcentreerd, herhaaldelijk lezen van een tekst, gericht op het redeneren over, analyseren van en reflecteren op de tekstinhoud. Daarbij sloten we aan bij de principes van observerend leren (bijv. Bandura, 1986; Braaksma, 2002), bij een meer flexibele opvatting van strategie-instructie (bijv. NAEP Reading Framework, 2017), bij recente inzichten over de functionele inbedding van tekstbegrip (bijv. McCrudden et al., 2011), en bij het belang van het kunnen bepalen van de bruikbaarheid en betrouwbaarheid van informatie, zoals dat vooral is uitgewerkt in theorie rondom meervoudig-tekstbegrip (bijv. Braasch et al., 2018). Daarnaast namen we het recente inzicht mee dat vaardige lezers zich niet alleen van minder vaardige lezers onderscheiden in hun opgebouwde tekstbegrip, maar ook in hun taakmodel, in hun planning en aanpak van een leestaak (Britt et al., 2018).

Vanuit deze uitgangspunten ontwikkelden we een lessenserie die in veel opzichten als een tegenhanger kan gelden van de courante tekst-met-vragen-aanpak. In veel observerend-leren-studies wordt leren door te observeren doorgaans afgewisseld met leren door te doen (bijv. Elving-Heida, 2019). De lessenserie wisselde daarom steeds af tussen enerzijds observerend-leren-lessen, waarin leerlingen nadenken over (de opbouw van) tekstbegrip, en anderzijds werklessen, waarin ze het eerder geleerde in de praktijk kunnen toepassen. In de observerend-leren-lessen beschrijven leerlingen het leesproces van leeftijdsgenoten en van henzelf, ze reflecteren erop en evalueren het, waarmee de effectiviteit van hun eigen leesproces (en daarmee hun inzet van leesstrategieën) vanzelfsprekend gespreksonderwerp wordt. In alle lessen werd bovendien het lezen van teksten aangeboden binnen een betekenisvolle context, met een duidelijk leesdoel, zodat leerlingen kunnen reflecteren op de aansluiting van teksten op dit leesdoel, met andere woorden: ze staan stil bij de betrouwbaarheid en bruikbaarheid van de teksten. In de werklessen sloten de leestaken steeds aan bij dit vooraf-gegeven doel, leestaken wisselden in vorm sterk af en moesten worden verricht zonder dat de vooraf-bestudeerde tekst nog raadpleegbaar was; de vervolgtask werd zo een controlemiddel op het vooraf opgebouwde begrip, waarmee het belang van een zelfstandige vooraf-tekstbestudering werd benadrukt. In de laatste twee lessen van de lessenserie uit hoofdstuk 6 stond bovendien ook het taakmodel van leerlingen ter discussie: hoe moet een complexe leestaak uitgevoerd worden, wil men een pittige schoolse leestaak afdoende hebben afgerond?

In hoofdstuk 5 en 6 zagen we dat deze alternatieve benadering over het algemeen effectief bleek. Leerlingen gaven in beide hoofdstukken aan dat ze ná de lessenserie

zelf veel vaker leesgedrag zijn gaan inzetten dat kenmerkend is voor diep lezen: ze herlezen na de lessenserie naar eigen zeggen bijvoorbeeld vaker alinea's en vatten een tekst vaker tussentijds in hun hoofd samen. Deelnemende leraren en leerlingen reageerden daarnaast over het algemeen positief op de lessenserie ("tof", "een eye-opener"). In hoofdstuk 5 had de groep leerlingen daarnaast blijkens de afgenomen begripstoetsen ná de experimentele lessenserie meer tekstbegrip opgebouwd dan de groep leerlingen die reguliere leeslessen had gevolgd. In hoofdstuk 6 zagen we dat de experimentele lessenserie een effect op het leren had onder twee voorwaarden. Ten eerste, leerlingen moesten wel zorgvuldig de lessenserie doorwerken; hun 'wijze van uitvoering' bepaalt in beide onderzoekscondities in hoeverre ze wat van de lessenserie opsteken: leerlingen die bijvoorbeeld nauwelijks of niet de moeite namen hun antwoorden te controleren, leren nauwelijks of niet van de lessenserie. Ten tweede, de docent voert klassengesprekken in de observerend-leren-lessen en is als regisseur in de les aanwezig. Alleen in de onderzoeksconditie 'Docent-gestuurd leren' kunnen we een algemeen effect op het tekstbegrip van leerlingen aantonen. Als de docent de voortgang niet regisseert en de reflectie in de observerend-leren-lessen niet klassikaal leidt, dan wordt het individuele engagement van leerlingen bepalend of en in welke mate ze wat van de lessenserie leren.

We concludeerden daarom dat werkvormen zónder een actieve rol van de docent bij lessenseries zoals deze behoedzaam moeten worden ingezet.

3 Implicaties voor de wetenschap

Zoals we in de inleiding al benadrukten, is het onderzoek in dit boek vakdidactisch en -inhoudelijk sterk verankerd in de bestaande praktijk in het schoolvak Nederlands, en dan met name in de vwo bovenbouw – in de volgende paragraaf, § 4, zullen we bij de implicaties van ons onderzoek voor het schoolvak Nederlands nog uitgebreid stilstaan. Eerst staan we echter stil bij de wetenschappelijke implicaties van ons onderzoek. In § 3.1 werken we allereerst enkele methodologische kanttekeningen bij ons onderzoek uit. In § 3.2 gaan we vervolgens in op de wetenschappelijke implicaties van ons onderzoek en geven we suggesties voor mogelijk vervolgonderzoek.

3.1 Methodologische kanttekeningen

In deze paragraaf gaan we in op vier kanttekeningen bij de onderzoeksmethodologie in dit boek. We bespreken achtereenvolgens: a) sterkten en zwakten van oogbewegingsonderzoek, b) sterkten en zwakten van de *mixed-method*-benadering, c) de meetproblematiek, en d) de generalisatieproblematiek.

Sterkten en zwakten van oogbewegingsonderzoek. In hoofdstuk 2 en 3 maakten we primair gebruik van oogbewegingsonderzoek. Eye-tracking vormt vertrekpunt van veel Nederlands onderzoek naar tekstbegrip (Cozijn, 2000; Kraal, 2020; De Leeuw,

2015; Van der Schoot et al., 2012; Van Silfhout, 2014; etc.) en van veel internationaal onderzoek naar tekstbegrip (vgl. de overzichten in Holmqvist & Andersson, 2017; Rayner, 1998, 2009). Eye-tracking stond ons in hoofdstuk 2 en 3 toe op relatief non-invasieve wijze het vooraf-lees- en antwoordgedrag van vwo-leerlingen feitelijk te beschrijven bij vier sets teksten met vragen; een hardop-denkonderzoek met eenzelfde opzet zou al snel te omvattend en ingrijpend voor de deelnemers en bovendien uiterst arbeidsintensief voor de onderzoeker zijn geweest.

Daarbij had oogbewegingsonderzoek belangrijke voordelen én beperkingen. Vooropgesteld: het stond ons relatief fijnmazige observaties toe. Zowel in hoofdstuk 2 als in hoofdstuk 3 maakten we gebruik van het onderscheid in *first-pass*- en *second-pass*-leesmaten, met name om te kunnen bepalen in hoeverre vwo-leerlingen hoofdzaken (kernzinnen) van bijzaken (niet-kernzinnen) onderscheidde, een noodzakelijke voorwaarde voor een goed begrip van een tekst (Hyönä & Niemi, 1990; Kintsch, 1998; Lorch et al., 1985; Yeari et al., 2015). Bij de *first-pass*-maat, die met name lagere-orde-begripsdimensies in kaart brengt (Holmqvist & Andersson, 2017; Kraal, 2020; Kuperman & Van Dyke, 2011), vonden we in hoofdstuk 2 dat vwo 4-leerlingen wat meer leestijd besteden aan kernzinnen dan aan niet-kernzinnen; bij de *second-pass*-maat, die met name hogere-orde-begripsdimensies in kaart brengt (ibidem), konden we in hoofdstuk 2 een dergelijke positieve kernzindiscriminatie niet aantonen. In hoofdstuk 3, waarbij we over aanzienlijk meer deelnemers en aanzienlijk meer *power* beschikten dan in hoofdstuk 2, vonden we een meer genuanceerde bevinding. Daar konden we een redelijk verschil in leestijd van kernzinnen en niet-kernzinnen aantonen in de *first-pass*-maat en een nipt verschil in de *second-pass*-maat: in eerste instantie besteden vwo-leerlingen meer leestijd aan kernzinnen dan aan niet-kernzinnen, maar dit onderscheid verdwijnt vrijwel bij herlezing. Deze tamelijk fijnmazige bevinding uit hoofdstuk 3 sluit aan bij de bevindingen in Yeari et al. (2015): als lezers weten dat ze na tekstlezing begripsvragen over de *gehele* tekst moeten beantwoorden, detecteren ze bij eerste lezing weliswaar kernzinnen, maar bij herlezing richten ze zich op de gehele tekst. Eye-tracking heeft als belangrijk voordeel dat het de tekstverwerking door lezers in het moment zelf relatief non-invasief registreert; hardop-denkonderzoek had dergelijk fijnmazige, genuanceerd afwijkende leespatronen als deze niet of nauwelijks kunnen blootleggen, zeker aangezien de deelnemers in hoofdstuk 2 en 3 zich relatief zelden tot herlezing van tekstdelen bleken te zetten.

Tegelijk liepen we aan tegen bekende beperkingen van deze onderzoeksmethode: eye-tracking kan weliswaar nauwkeurig verschillen in kwantiteit aan aandacht in kaart brengen, maar niet verschillen in kwaliteit aan aandacht; bovendien is het vaak moeilijk een eenduidige interpretatie toe te kennen aan eye-track-resultaten (Bax, 2013; Cozijn, 2000; Edmonds & Kennedy, 2016; Miller, 2015). In hoofdstuk 3 bijvoorbeeld bleken verschillen tussen vwo 4- en vwo 6-leerlingen meermaals niet significant, en verschillen die wél significant waren, spoorden soms in het geheel niet met onze theoretische verwachtingen. Zo blikten vwo 4-leerlingen tijdens vooraf-lezen vaker terug naar eerdere tekstdelen dan vwo 2-leerlingen, maar ook vaker dan vwo 6-leerlingen. Lazen de vwo 4-leerlingen de teksten misschien met meer

aandacht en scherpte dan de vwo 6-leerlingen en herlezen ze daarom vaker tekstdelen? Zagen we hier een relatief vaardige steekproef vwo 4-leerlingen versus een wat minder vaardige steekproef vwo 6-leerlingen? Of was het juist andersom: hadden de vwo 4-leerlingen meer behoefte aan herlezing dan de vwo 6-leerlingen, omdat ze meer moeite met de teksten hadden? Pas door in de eye-track-data ook de resultaten van de retrospectieve leerlinginterviews te verdisconteren, werd aannemelijk dat de vwo 6-leerlingen de teksten tijdens het vooraf-lezen wel degelijk beter hadden begrepen, ondanks dat ze minder hadden teruggekeken. Waarom de vwo 4-leerlingen vaker tekstdelen teruglezen dan vwo 6-leerlingen, blijft daarmee een open vraag.

Bij sterk theoretisch gedreven beschrijvend onderzoek (naar bijvoorbeeld het lezen van kernzinnen of structuurmarkeringen) vormt oogbewegingsonderzoek een krachtige onderzoeksmethode (vgl. bijv. Hyönä & Niemi, 1990; Van Silfhout, 2014); in meer ‘waardevrij’ beschrijvend onderzoek zoals dat in hoofdstuk 2 en 3, is het raadzaam een *mixed-method*-benadering te verkiezen, die lees- en/of antwoordgedrag (deels) kwalitatief kan benaderen, zoals in Bax (2013) en Brunfaut & McCray (2015) – studies die inspiratiebron vormden voor ons onderzoek in hoofdstuk 2 en 3. We gaan op het belang van deze onderzoeksmethode hieronder nader in.

Sterkten en zwakten van de mixed-method-benadering. In alle studies in dit boek neemt de kwantitatieve benadering een belangrijke plaats in: in hoofdstuk 2 tot en met 6 speelt multiniveau-modellering in de analyse van de data een hoofdrol (zie Kreft & De Leeuw, 2007; Quené & Van den Bergh, 2004). Echter, zowel in hoofdstuk 2 en 3 als in hoofdstuk 6, maakten we daarbij gebruik van een *mixed-method*-benadering die kwantitatief en kwalitatief onderzoek combineert (Creswell & Plano Clark, 2018; Edmonds & Kennedy, 2016).

Deze benadering kent in ons onderzoek belangrijke voordelen. Nemen we bijvoorbeeld hoofdstuk 3, waarin gebruik werd gemaakt van een *explanatory sequential* design (Creswell & Plano Clark, 2018, p. 60): ná het eye-track-experiment volgde een interview met deelnemers waarmee we het rationale achter de geregistreerde oogbewegingen konden verklaren. In eerder onderzoek naar de tekst met vragen werd de complexe interactie tussen tekst en vraag nog weleens versimpeld. Zo zijn de vraagbestudering en tekstraadpleging naar aanleiding van de vraag herhaaldelijk in kaart gebracht (bijv. Bax, 2013; Brunfaut & McCray, 2015; Embretson & Wetzel, 1987; Rupp et al., 2006), maar de relatie tussen vooraf-tekstbestudering enerzijds en de vraagbeantwoording en het antwoord anderzijds bleef onderbelicht. Door een retrospectief interview in te bouwen over het rationale achter de vraagbeantwoording, konden we deze relatie in meer detail blootleggen. De studie in hoofdstuk 3 is daarmee bij ons weten de eerste studie die het vooraf-lees- en antwoordproces bij de tekst met vragen holistisch heeft benaderd: de relaties tussen de deelprocessen (‘vooraf-lezen’, ‘vraag lezen’, ‘vraag beantwoorden’) en het uiteindelijke product (‘antwoord op de vraag’) brachten we voor vier teksten met vragen in kaart. Daarmee is het bij ons weten ook de eerste studie die vanuit deze onderzoeksbenadering cross-sectionele verschillen tussen jongere en oudere leerlingen in kaart brengt.

Ook in hoofdstuk 5 en 6 speelde de *mixed-method*-benadering een belangrijke rol. Daar maakten we gebruik van een *convergent design* (Creswell & Plano Clark, 2018): vooral in hoofdstuk 6 trianguleerden we gelijktijdig verzamelde kwantitatieve en kwalitatieve data om zicht te krijgen op de implementatie, de receptie en het effect van de experimentele lessenserie over diep lezen – een werkwijze zoals die ook is te vinden in bijvoorbeeld de effectstudies in Bouwer en Koster (2015) en Elving-Heida (2019). Dit convergente onderzoeksontwerp vormde vooral een belangrijk instrument om de implementatietrouw in kaart te brengen (Carroll et al., 2007): we gebruikten lesobservaties, logboeken van docenten en de leerlingantwoorden op opgaven om zicht te krijgen op de uitvoering van de lessenserie. Echter, ook hier hadden we idealiter ons op nog meer gegevens kunnen baseren: zo namen we in het beperkte aantal lesobservaties niet de kwaliteit van de gevoerde klassengesprekken mee, noch namen we in onze bestudering van de leerlingantwoorden bij de lessenserie (uitgebreid) de kwaliteit van de beantwoording mee – hoewel een dergelijk onderzoek interessante perspectieven op de werkzaamheid van de lessenserie had kunnen openen. Het arbeidsintensieve karakter van dergelijk onderzoek bij een relatief grote steekproef als het onderzoek in hoofdstuk 6 verhinderde dit.

Graag onderstrepen we hier de methodologische meerwaarde van *mixed-method*-benaderingen in beschrijvend onderzoek en effectonderzoek voor leesonderwijs.

De meetproblematiek. De meetproblematiek speelt in dit boek een bijzondere rol. Dat geldt bij uitstek voor hoofdstuk 4, waarin het functioneren van een tekstbegrips-*oefening* is onderzocht aan de hand van tekstbegripstoetsen, in een toetssetting. Het is verleidelijk de uitkomsten uit dit onderzoek over de lees*oefening* zonder enig voorbehoud óók op te vatten als uitkomsten over de lees*toets*. Daarmee zou de studie uit dit hoofdstuk dan eerder onderzoek kunnen onderbouwen waarin de validiteit van de tekst met vragen als leestoets ter discussie is gesteld, zoals dat van Kamalski (2007). Het is echter goed hier te benadrukken dat in dit boek toetsing van tekstbegrip zelf geen direct onderzoeksobject is geweest: het is vanuit hoofdstuk 4 aannemelijk dat een tekst-met-vragen-didactiek leerlingen niet aanzet tot een diepgaande zelfstandige tekstbestudering, maar daarmee is niet gezegd dat een tekst met vragen tekstbegrip niet valide zou kunnen meten. Desalniettemin denken we dat het onderzoek *indirect* wel implicaties heeft voor toetsing: in § 4.2 zullen we nader op deze kwestie ingaan.

Het gekozen onderzoeksdesign heeft natuurlijk een belangrijke invloed op de mate waarin men het effect van een interventie valide kan meten (Shadish et al., 2002). Zo bleek in hoofdstuk 5 nogal wat onderzoek naar EMME's niet te voldoen aan de basale voorwaarde van een pretest en posttest (zie het overzicht in Salmerón et al., 2020); aan die voorwaarde voldoen de onderzoeken in hoofdstuk 5 en 6 wel, waarbij we in beide studies de effectiviteit van de lessenserie konden aantonen wanneer tekstbegrip naderhand werd gemeten zowel via een cloze-taak als via een reguliere tekst met begripsvragen. En waar we in hoofdstuk 5 het experiment nog afnamen op één middelbare school met actieve hulp van de onderzoeker, namen

in hoofdstuk 6 de docenten Nederlands op zeven middelbare scholen het experiment geheel zelfstandig af. Daarbij maakten we in dat laatste hoofdstuk bovendien gebruik van een *switching-replications*-design waarmee we de effectiviteit van de interventie tweemaal konden aantonen.

Relevant bij de onderzoeken in hoofdstuk 5 en 6 is naar ons idee vooral de vraag wélk meetinstrument ingezet had moeten worden: gelet op de uitkomsten in met name hoofdstuk 4, zou men in hoofdstuk 5 en 6 kunnen spreken van een gebrekkige *test alignment* met de leerdoelen van de lessenserie (Biggs, 1996; Van Silfhout, In voorbereiding.) In de experimentele lessenserie in deze hoofdstukken werd namelijk in instructie en taakinrichting het belang benadrukt van een zelfstandige bestudering van de tekst; inhoudelijk sterk wisselende vervolgtaken werden aangeboden waarbij leerlingen telkens niet langer de bestudeerde teksten konden raadplegen. In de eindmeting werd echter deels teruggevallen op gaten teksten én op de klassieke begripsvragentoets waarbij leerlingen bij de vragen de teksten wél konden raadplegen. Vooral deze laatste toets was volgens enkele deelnemende docenten onvoldoende in lijn met de leerdoelen: leerlingen hadden verwacht dat ze in de toets opgaven zouden moeten maken zonder dat ze de tekst nog konden raadplegen – een zinnige observatie! De vraag is in hoeverre de reguliere tekst-met-vragen-toets de meerwaarde van onze meer procesgerichte benadering heeft kunnen meten; ze lijkt immers vooral om de beantwoording van vragen te draaien, blijkt ons onderzoek in hoofdstuk 2 tot en met 4; de intensiteit van een zelfstandige bestudering meet ze (waarschijnlijk) niet of nauwelijks.

Rationale achter de gekozen benadering was echter dat we binnen de courante tekstbegripstoetsing in het schoolvak Nederlands een positief effect van de lessenserie wilden aantonen. Leerlingen in het Nederlandse onderwijs bleken weinig tot geen ervaring te hebben met leestoetsvormen waarbij de tekst bij de begripsvragen niet langer raadpleegbaar blijft; bij dergelijke begripstoetsen zou een hogere antwoordscore op na de lessenserie vooral ook een bekendheid van leerlingen met een afwijkende toetsvorm hebben gemeten. Echter, uitgaande van het onderzoeksresultaat uit met name hoofdstuk 4 – namelijk dat een diepgaande voorafbestudering van een tekst waarschijnlijk niet leidt tot meer goede antwoorden op begripsvragen – betekent deze gebrekkige *test alignment* waarschijnlijk dat de gemeten leerwinst in tekstbegrip in hoofdstuk 5 en 6 tot op onbekende hoogte een onderschatting vormt van het werkelijke leereffect van de lessenserie.

Een andere kwestie die we hier willen noemen, is dat we in hoofdstuk 2 tot en met 4 enkel begripsvragen op het niveau van het situatiemodel hebben onderzocht, waarmee we goeddeels aansloten bij de staande praktijk in het schoolvak Nederlands. We hebben in ons onderzoek geen onderzoek gedaan naar tekstonafhankelijke bevragingssystemen (zoals deels voorgestaan in bijv. Hendrix & Hulshof, 1994), noch naar een bevragingssysteem dat het domein evalueren en reflecteren geheel centraal stelt (bijv. een uitgebreide set vragen over de betrouwbaarheid en bruikbaarheid van tekstuele informatie). En weliswaar hebben andersoortige taken zoals de sorteertaak en de mental model-taak een plaats gekregen in onze experimentele lessenserie in hoofdstuk 5 en 6, we hebben het functioneren van deze

taken als oefening of als toets niet nader onderzocht. Dergelijk vervolgonderzoek voor het schoolvak Nederlands is overigens zeker wenselijk (vgl. Kamalski, 2007; Van der Meulen, 2015; Robben, 2012; Vermeij, 2011).

Generalisatieproblematiek. Zoals we al in de inleiding aangaven, is de mogelijkheid tot generalisatie van de onderzoeksuitkomsten impliciet een van de schragende uitgangspunten in ons onderzoek. Ze heeft zich in dit boek geuit in:

- 1) een natuurlijke onderzoeksomgeving: in hoofdstuk 2 en 3 namen we het eye-track-onderzoek af op scholen (en dus niet in een universitair laboratorium), en in hoofdstuk 4 tot en met 6 vormde het klaslokaal Nederlands plaats van handeling;
- 2) in een statistische hoofdrol voor multiniveau-modellen, die ons toestaan te generaliseren over teksten, taken, begripsvragen, klassen en/of leerlingen;
- 3) in een werkwijze waarbij we tekstbegrip steeds operationaliseerden aan de hand van meerdere teksten: in hoofdstuk 2 en 3 gebruikten we vier teksten, in hoofdstuk 4 gebruikten we twee teksten, en in hoofdstuk 5 en 6 gebruikten we voor de natoets twee teksten;
- 4) in relatief grootschalige onderzoeksopzetten: de steekproef ($N = 52$) uit hoofdstuk 3 is aanmerkelijk groter dan die in vergelijkbaar *mixed-method*-oogbewegingsonderzoek naar lees- en antwoordgedrag (vgl. Bax, 2013; Brunfaut & McCray, 2015), in hoofdstuk 4 waren 26 vwo 5-klassen ($N = 572$) betrokken en in hoofdstuk 6 namen 27 vwo 4-klassen ($N = 701$) de lessenserie af; al met al waren daarmee 1456 vwo-leerlingen van dertien verschillende scholen in Nederland in ons onderzoek betrokken;
- 5) in een replicatie van onderzoeksresultaten over hoofdstukken heen; replicatie kan consistentie in onderzoeksresultaten aantonen, geen onbelangrijk actueel aandachtspunt voor wetenschappelijk onderzoek (Open Science Collaboration, 2015). De resultaten uit hoofdstuk 2 en 3 worden ondersteund en verklaard door de bevindingen in hoofdstuk 4, en de onderzoeksbevindingen in hoofdstuk 5 worden in grote lijnen gerepliceerd in hoofdstuk 6. En in hoofdstuk 6 maakten we gebruik van een *switching-replications*-onderzoeksontwerp, waarmee we binnen het experiment tweemaal het effect van de interventie konden aantonen.

Generalisatie blijft evenwel in dit boek – zoals in veel onderwijsonderzoek – een probleem. Zo geldt voor hoofdstuk 2 en 3 dat we alleen het vooraf-lees- en antwoordgedrag van een random steekproef vwo-leerlingen op één middelbare school in het zuiden van Nederland in beeld hebben gebracht. Daarbij geldt weliswaar dat de resultaten in deze twee hoofdstukken in veel opzichten onderbouwd en verklaard kunnen worden door de studie in hoofdstuk 4, echter, in hoeverre de studies in hoofdstuk 2 en 3 generaliseerbaar zijn over scholen, hangt deels af van de herkenbaarheid van de resultaten. De resultaten sloten in elk geval, grotendeels onverwacht, sterk aan bij veelgehoorde klachten over de tekst met vragen in het schoolvak Nederlands dat leerlingen er niet zelfstandig mee leren lezen en vooral er vragen mee leren beantwoorden.

Voor het onderzoek in hoofdstuk 5 geldt verder dat het slechts afgenomen werd op één middelbare school en dat de onderzoeker een belangrijke actieve rol speelde in het afnemen van de lessenserie, naast de docent Nederlands van de deelnemende vwo 3- en vwo 4-klassen. Hierbij moet dan wel worden opgemerkt dat we de resultaten uit hoofdstuk 5 konden repliceren in hoofdstuk 6, waar de lessenserie in een bijgestelde versie werd afgenomen op zeven andere middelbare scholen.

Andere generalisatiekwesaties dienen hier nogmaals benadrukt. Zo heeft het onderzoek in dit boek in strikte zin alleen betrekking op vaardige lezers, vwo-leerlingen, en dan met name bovenbouwleerlingen. In hoeverre de resultaten kunnen gelden voor lezers uit andere schoolsoorten, is een lastig te beantwoorden vraag. We gingen er al eerder op in dat er ons geen reden lijkt om de belangrijkste bevinding uit hoofdstuk 2 tot en met 4 ('Pas op met de tekst met vragen!') niet evenzeer voor havo- of vmbo-leerlingen te laten gelden. Desalniettemin zijn er tal van open vragen. In hoeverre gedragen vmbo- of havo-leerlingen zich anders bij het vooraflezen van teksten en het beantwoorden van vragen dan vwo-leerlingen? En in hoeverre is de lessenserie uit hoofdstuk 6 ook werkzaam bij vmbo- en havo-leerlingen?

Daarnaast geldt dat we binnen de vwo-groepen geen onderscheid hebben gemaakt tussen vaardige en minder vaardige lezers. Weliswaar vormen vwo-leerlingen binnen het Nederlandse voortgezet onderwijs de vaardige groep lezers (Expertgroep Doorlopende Leerlijnen Taal, 2008), de verschillen in leesvaardigheid tussen vwo-leerlingen kunnen desalniettemin zeer groot zijn (vgl. Gubbels et al., 2019; Van Silfhout, 2014, p. 153). In hoeverre de interventies uit hoofdstuk 5 en 6 meer of minder werkzaam zijn voor enerzijds relatief vaardige vwo-leerlingen en anderzijds relatief minder vaardige vwo-leerlingen, is een open vraag. Ons onderzoek had enkel betrekking op groepen vwo-leerlingen uit verschillende jaarlagen.

Ten slotte heeft het onderzoek steeds alleen betrekking op relatief korte uiteenzettende of betogende teksten uit Nederlandse dagbladen en tijdschriften, zoals die veelal in huidige leergangen Nederlands te vinden zijn (bijv. Frank et al., 2019), evenals in toetsing rondom het schoolvak Nederlands (bijv. CvTE, 2022a, 2022b). Daarmee is tekstbegrip ernstig versmald – wat echter vooral als een kritiek op de leeslessen in het huidige schoolvak Nederlands kan worden beschouwd: de staande praktijk in de leesles, zoals vormgegeven in huidige leergangen Nederlands, vormde immers uitgangspunt van dit boek.

3.2 Suggesties voor vervolgonderzoek

In een aantal opzichten vraagt het onderzoek in dit boek, denken we, om vervolgonderzoek. We lopen vier richtingen voor vervolgonderzoek in onderstaande langs, waarbij we enkele malen teruggrijpen naar opmerkingen uit de voorgaande paragraaf. We zullen achtereenvolgens de wenselijkheid van vervolgonderzoek bespreken naar 1) oplossingsprocedures bij leestaken; 2) de tekst met vragen als leestoets; 3) voorwaarden voor de effectiviteit van observerend leren; en 4) de effectiviteit van soorten modellen bij observerend leren.

Onderzoek naar oplossingsprocedures bij leestaken. In hoofdstuk 4 gingen we in op het taakmodel van vaardige lezers bij een tekst met vragen, zoals vaak aangeboden in de context van het schoolvak Nederlands. Daarbij vertrokken we vanuit het RESOLV-model van Britt et al. (2018) dat leesgedrag bij complexe taken wil voorspellen en verklaren. We maakten van daaruit aannemelijk dat de tekstraadpleging tijdens de antwoordfase bij complexe situatiemodelvragen een groot effect zal hebben op het uiteindelijke antwoord; in hoeverre de intensiteit van vooraf-lezen een effect heeft, konden we vanuit dit model niet eenduidig voorspellen. Uit ons experimentele onderzoek bleek vervolgens dat voor complexe situatiemodelvragen, in een situatie waarin de tekst na vooraf-bestudering raadpleegbaar bleef, de intensiteit van het vooraf-lezen geen aantoonbaar effect heeft op de antwoordscore; de tekstraadpleegbaarheid tijdens het antwoordproces had daarentegen inderdaad een groot effect op de antwoordscore. Met deze bevinding kan het model van Britt et al. (2018) meer scherpte worden gegeven.

We vonden met de studie uit hoofdstuk 4 daarnaast onderbouwing voor twee belangrijke hypothesen die Britt et al. (2018) aan hun model ontleen. Ten eerste hun *minimal task elaboration hypothesis*: lezers plannen vooraf niet uitvoerig hoe ze een leestaak aanpakken. Lezers lezen, blijktens onze studie, inderdaad bij een tekst-met-vragen-toets de tekst vooraf enkel oppervlakkig door; pas bij de vragen gaan ze over op een nauwkeurige, uitgebreide werkwijze. En ten tweede hun *parsing the context hypothesis*: vanuit de taakcontext bepalen lezers in welke mate ze zich voor een leestaak inzetten. Het ene experiment namen we in onze studie af als summatieve toets in een toetsweek, terwijl we het andere experiment als diagnostische toets afnamen in een les Nederlands. Ondanks het feit dat één conditie uit het ene experiment inhoudelijk en procedureel overlapt met één conditie uit het andere experiment, bleken er, waarschijnlijk vanuit deze afwijkende taakcontext, grote verschillen te bestaan in antwoordscore tussen deze twee condities: de leerlingen die de summatieve toets maakten, scoorden gemiddeld aanmerkelijk hoger dan de leerlingen die de diagnostische toets maakten. Vooral bij de pittige tekst-met-vragen viel dit verschil sterk op. Al met al bleek het model van Britt et al. (2018) voor onze studies zeer bruikbaar.

Er is nog veel ruimte voor vervolgonderzoek naar het functioneren van deze leestaak én andere leestaken, waarvoor het model van Britt et al. (2018) uitgangspunt kan zijn. In hoofdstuk 4 bespraken we welke begrips- en vormkenmerken van een begripsvraag binnen de huidige tekst-met-vragen-aanpak theoretisch bepalen of en in welke mate een vaardige lezer na vooraf-lezen de tekst zal raadplegen voor de beantwoording van een vraag. Zo zijn er bepalende begripskenmerken van de vraag: Moeten leerlingen voor de beantwoording vooral op hun tekstbasis of op hun situatiemodel een beroep doen? Moeten ze een (lokale) microbewering of een (alinea-overstijgende) macrobewering in de tekst isoleren? En is deze bewering relatief eenvoudig en expliciet terug te vinden of is ze juist complex en impliciet? Daarnaast wezen we op vormkenmerken: is een vraag open (waarbij een leerling het antwoord zelf moet produceren) of juist gesloten (waarbij een leerling het juiste antwoord enkel hoeft te selecteren)? Andere vormkenmerken noemden we daar niet – maar er

zijn er natuurlijk meer, zoals het zoekbereik van de vraag (antwoord zoeken binnen één alinea, binnen de hele tekst of zelfs tussen meerdere teksten) en de talige complexiteit van de vraagstelling (vgl. Cerdán et al., 2019; Mañez et al., 2019). Al met al geldt: een eenvoudige gesloten vraag op tekstbasisoniveau zal minder snel tot uitgebreide tekstraadpleging leiden dan een complexe open vraag op situatiemodelniveau – zoals in de experimenten in de hoofdstukken 2 tot en met 4.

Er blijven echter veel open vragen waarop het theoretisch model van Britt et al. (2018) toetsbare antwoorden biedt. Vaardige lezers zullen de tekst vooraf doorgaans lezen, omdat dit het vraagbeantwoordingsproces vergemakkelijkt en hen zelfs incidenteel toestaat direct de vraag te beantwoorden, zo zagen we in hoofdstuk 4. Geldt deze keuze tot vooraf-lezen echter ook voor relatief minder vaardige lezers, zoals vmbo-leerlingen, of slaan deze leerlingen juist doorgaans het vooraf-lezen van de tekst over (zoals docenten Nederlands weleens aangeven)? En is de soort of lengte van de tekst hierop van invloed? In het onderzoek uit hoofdstuk 4 zetten de relatief complexe situatiemodelvragen leerlingen dikwijls aan tot een zeer omstandig antwoordproces; geldt ditzelfde voor relatief eenvoudige lokale tekstbasisvragen? In het onderzoek uit hoofdstuk 4 vonden we dat het intensief vooraf-lezen van de tekst geen aantoonbaar effect heeft op de uiteindelijke correctheid van het antwoord; in hoeverre geldt dit echter ook voor andere leestaken zoals de sorteertaak en de samenvatting? Vervolgonderzoek naar deze en andere vragen lijkt zeer wenselijk.

Onderzoek naar de tekst met vragen als leestoets. We stelden in de vorige paragraaf al vast dat het onderzoek in dit boek enkel het functioneren van de tekst met vragen als oefening in de lessen tekstbegrip onderzocht, met name door deze als leestoets aan te bieden in hoofdstuk 4; we onderzochten dus niet het functioneren van de leestoets zelf. Vanuit dit onderzoek stelden we vast – we bespraken het al in § 3.1 in dit hoofdstuk – dat men in het algemeen terughoudend lijkt te moeten zijn met de inzet van deze oefening in het leesonderwijs. Desalniettemin sluiten de bevindingen in met name hoofdstuk 4 aan bij recente twijfels over de constructvaliditeit van de tekst-met-vragen-toets (met name O'Reilly et al., 2018).

De afgelopen 25 jaar is steeds meer onderkend dat de opbouw van tekstbegrip tijdens het lezen wordt beïnvloed door het doel dat lezers zichzelf met lezen stellen (vgl. Gil et al., 2010; Linderholm & Van den Broek, 2002; Yeari et al., 2015; etc.): lezers die een tekst ter ontspanning willen lezen, lezen deze anders dan wanneer ze deze voor hun studie willen bestuderen. In het model van Britt et al. (2018) worden de implicaties van dit doelmatig lezen voor tal van leestaken consequent doordacht: welke impact heeft een specifieke leestaak op de constructie van taaken contextmodellen van lezers? Een belangrijke kritiek op de reguliere tekst-met-vragen-leestoets vanuit deze theorievorming lijkt dat lezers erbij vooraf hooguit een generiek, vaag leesdoel krijgen aangereikt ('Bestudeer de tekst en beantwoord erna de vragen'); pas bij de vragen weten leerlingen waarop ze in een tekst moeten letten. Het impliciete doel van deze toets is voor lezers van meet af aan het goed kunnen beantwoorden van de vragen, en vooral hierop zullen ze pragmatisch hun taak- en contextmodellen inrichten.

Daarmee samenhangend is een belangrijke kritiek dat lezers met de tekst met vragen niet noodzakelijkerwijs een coherent, adequaat situatiemodel van de *totale* tekstinhoud hoeven op te bouwen, noch tijdens vooraf-lezen, noch tijdens het beantwoorden van de vragen. De meeste leestheorieën gaan er echter van uit dat lezers tot een adequaat begrip van een tekst komen als ze een coherent tekstconform situatiemodel ervan hebben opgebouwd (vgl. Kintsch, 1998; Magliano et al., 2018; Van den Broek, 2012; Van den Broek et al., 1999), waarop toetsing van diep begrip idealiter dus aansluit (bijv. Magliano et al., 2018). Aan deze voorwaarde voldoet de reguliere tekst-met-vragen-toets hoogstwaarschijnlijk slechts in gebrekkige mate.

Zo'n coherent diepgaand situatiemodel hoeven leerlingen – allereerst – namelijk niet op te bouwen tijdens vooraf-lezen, zo suggereert ons onderzoek uit hoofdstuk 4. Doordat lezers uit ervaring weten dat ze de tekst bij het beantwoorden van de vragen kunnen raadplegen, gebruiken ze het vooraf-lezen vooral om hun impliciete doel te bereiken: het goed beantwoorden van de vragen. Op welke delen of dimensies in de tekst de vragen zullen ingaan, kunnen lezers tijdens vooraf-lezen moeilijk voorspellen; de vragen vormen immers voor hen vaak een lukrake verzameling, een 'grabbelton' (vgl. Pearson & Hamm, 2005). Bij het vooraf-lezen richten lezers zich daarom vooral op de gehele tekstbasis, met aandacht voor zowel hoofdzaken als bijzaken, zodat ze een overzicht hebben van wat waar in de tekst staat, en ze bouwen vooral een rudimentair situatiemodel op (vgl. Yeari et al., 2015).

Ook tijdens de beantwoordingfase hoeven leerlingen zo'n coherent diepgaand situatiemodel van de gehele tekst meestal niet op te bouwen. Tijdens de beantwoording van een vraag bouwen lezers een adequaat situatiemodel op bij dat tekstdeel dat (naar hun idee althans) relevant is voor de beantwoording van de vraag; hiervoor raadplegen ze de tekst vaak uitgebreid, zoals we zagen in hoofdstuk 2 en 3 (vgl. Bax, 2013; Brunfaut & McCray, 2015; Cohen & Upton, 2006; Rupp et al., 2006; etc.). Gedurende het beantwoorden van de vragen bouwen leerlingen daardoor niet zozeer een samenhangend situatiemodel op van de gehele tekst, maar veeleer een verzameling situatiemodellen bij die tekstonderdelen die de vragen centraal hebben gesteld. De optelsom van die verzameling situatiemodellen kán natuurlijk sterk inhoudelijk 'correleren' met een samenhangend situatiemodel van de gehele tekst – maar dat lijkt grotendeels afhankelijk van de vragen die worden gesteld.

Voor de goede orde: toetsing van tekstbegrip kan natuurlijk ook lagere-orde-begripsvaardigheden bevragen en hoeft zich niet uitsluitend te richten op het diepere begrip van teksten (vgl. Alderson, 2005; Cain & Oakhill, 2000; Van den Broek, 2012). De vraag in hoeverre de begripsvragentoets voor de meting van lagere-orde-begripsvaardigheden geschikt is (de integratie van zinsdelen tot een bewering bijvoorbeeld), valt echter buiten de kaders van dit boek.

Louter bezien vanuit de kaders van dit boek, vanuit diep begrip, lijkt er zeker reden om te zoeken naar alternatieven voor de traditionele tekst-met-vragen-toets. De laatste tien jaar zijn er interessante alternatieve inrichtingen van deze leestoets verkend die de genoemde problemen trachten te omzeilen. Een veelbelovende toetsvorm in dit opzicht is *Scenario-Based Assessment* (SBA: Sabatini et al., 2018,

2020): lezers krijgen hierin vanuit een betekenisvol scenario een duidelijk realistisch en herkenbaar leesdoel aangereikt, voordat ze een tekst (of teksten) moeten bestuderen; ze krijgen vervolgens louter opgaven aangeboden die geheel in lijn zijn met scenario en doel. Deze toetsvorm heeft inmiddels ingang gevonden in internationaal vergelijkende toetsing (vgl. OECD, 2019) en Amerikaanse voortgangstoetsing (vgl. Pearson et al., 2020). Ook in de Centraal Examens Nederlands havo/vwo zullen SBA's met ingang van 2024 hoogstwaarschijnlijk deels een plaats krijgen (vgl. Evers-Vermeul et al., 2022). SBA's hebben daarbij als voordeel dat ze relatief eenvoudig een plaats kunnen bieden aan meervoudig-tekstbegrip en het begrip van digitale media (Sabatini et al., 2020).

Een tweede, vooral voor enkelvoudig-tekstbegrip geschikte mogelijkheid is een toets met begripsvragen, zónder dat de tekst bij de vragen raadpleegbaar blijft. De laatste vijftien jaar is nauwkeurig de impact in kaart gebracht op vooraf-lees- en antwoordgedrag bij deze toets, als de tekst na bestudering niet langer raadpleegbaar is bij de vraagbeantwoording (Ferrer et al., 2017; Higgs et al., 2017; Schroeder, 2011; vgl. Johnston, 1984): de antwoordscore blijkt dan sterk afhankelijk van het vooraf-lezen. In ons onderzoek zijn we regelmatig teruggevallen op deze toetsvorm. Het lijkt ons een vanzelfsprekende, meer valide invulling dan de gebruikelijke vorm: in deze vorm geeft de antwoordscore namelijk wél vooral een indicatie in hoeverre lezers tijdens lezen een geslaagde, samenhangende mentale representatie van de tekstinhoud hebben weten op te bouwen (Schroeder, 2011). Vraag is dan wel in hoeverre deze inrichting steeds vertrekpunt kan vormen bij tekstbegripstoetsing: Wanneer worden teksten te lang voor lezers om te memoriseren en kan hun begrip niet meer zónder tekstraadpleging bevraagd worden? En zijn er mogelijkheden om ook meervoudig-tekstbegrip op deze wijze te bevragen?

Vervolgonderzoek zou scherper in kaart kunnen brengen welke gevolgen alternatieve inrichtingen van de begripsvragentoets zoals bovenstaande hebben op het vooraf-lezen, het antwoordgedrag en het uiteindelijke antwoord. De uitdaging voor de toekomst is toetsvormen te vinden en te onderzoeken die leesdoel, (vooraf-)tekstbestudering, verwerkingsopdracht en de uiteindelijke toetscore optimaal met elkaar in lijn brengen.

Onderzoek naar voorwaarden voor effectiviteit van observerend leren. In hoofdstuk 5 en 6 toonden we de effectiviteit aan van een lessenserie over diep lezen waarin observerend leren een belangrijk didactisch principe vormde. Daarmee schaarft het zich in een rij studies die de meerwaarde van observerend leren voor het schoolvak Nederlands laat zien, zowel voor het schrijfonderwijs (Bouwer et al., 2018b; Braaksma, 2002; Braaksma et al., 2004; Couzijn, 1995; Elving-Heida, 2019; Rijlaarsdam et al., 2005, 2008; etc.) als, meer incidenteel, voor het leesonderwijs (Couzijn, 1995; Keehnen et al., 2015).

Enkele studies naar observerend leren konden geen direct leereffect aantonen (bijv. Van der Loo, 2022) of slechts een zeer beperkt leereffect (bijv. Grenner et al., 2020; Groenendijk et al., 2011). In Van der Loo (2022) wordt daarom de vraag opgeworpen of cognitieve capaciteiten en andere leerlingkenmerken, samen met ken-

merken van de leeromgeving en van de *community of writers*, niet een dermate sterke impact op het leerproces hebben dat ze de werkzaamheid van observerend leren in de weg kunnen staan; vervolgonderzoek zou de oorzaken van deze wisselende effectiviteit van observerend leren dienen te verkennen. Gezien de soms wisselende resultaten in studies naar de effectiviteit van observerend leren lijkt dergelijk vervolgonderzoek ons inderdaad een goede zaak; tegelijk zouden we de vraagstelling daarbij willen preciseren.

Vooropgesteld: dat we in onze studies in hoofdstuk 5 en 6 positieve resultaten vonden, kan met tal van factoren samenhangen. In vergelijking met bijvoorbeeld Grenner et al. (2020) onderzochten we de effectiviteit van observerend leren immers bij een andere doelgroep (vwo 3- en vwo 4-leerlingen), een ander taaldomein (lezen) en een andere omgeving (middelbare scholen). Daarnaast dient nadrukkelijk vermeld dat we in deze hoofdstukken niet nader hebben onderzocht in hoeverre specifiek de observerend-leren-lessen in deze lessenserie bijdroegen aan de effectiviteit ervan; observerend leren vormt binnen onze lessenserie slechts één instrument in een breder didactisch instrumentarium, zij het een belangrijk instrument (vgl. Elving-Heida, 2019). Evenmin hebben we nader onderzocht in hoeverre vaardige en minder vaardige lezers verschillend van de lessenserie hebben geleerd, terwijl dat onderscheid in andere studies wel zinvol bleek (bijv. Braaksma, 2002).

Vervolgonderzoek zou wat ons betreft niet alleen moeten bezien onder welke omstandigheden, maar vooral ook onder welke voorwaarden observerend leren effectief is. Bandura (1986) stelt een viertal voorwaarden, willen leerlingen door observatie van anderen tot leren komen: er moet sprake zijn van: 1) aandacht voor het modelgedrag, 2) retentie: onthouden van het modelgedrag, 3) *motor production*: leerlingen moeten in staat zijn het gedrag van modellen te reproduceren, en 4) motivatie: leerlingen moeten het modelgedrag willen gaan vertonen. Braaksma (2002) en Braaksma et al. (2006) verbinden hieraan nog de voorwaarde van 5) evaluatie: leerlingen moeten niet alleen het gedrag van modellen beschrijven en erop reflecteren, maar het ook evalueren.

Twee vormen waarin we deze voorwaarden in onze experimentele lessenserie in hoofdstuk 5 en 6 hebben geïmplementeerd, verdienen dan extra aandacht. Ten eerste: we hebben de voorwaarden *aandacht* en *motivatie* trachten te versterken. We hadden reden eraan te twifelen of leerlingen bij onze interventies zonder extra stimulans wel voldoende aandacht en motivatie zouden opbrengen om de modellen te observeren. De vraag is namelijk in hoeverre leerlingen vanuit zichzelf hun eigen leesgedrag meestal als problematisch ervaren: de beoordeling ('kalibratie') van het eigen begrip door (jonge) lezers is vaak erg onbetrouwbaar (bijv. Glenberg & Epstein, 1987; Maki & Berry, 1984). Lezen lijkt over het algemeen ook als minder cognitief belastend ervaren te worden dan bijvoorbeeld schrijven (Kellogg, 1994). Mogelijk ervaren leerlingen bij leestaken minder behoefte aan voorbeeldgedrag. Daarom hebben we geprobeerd hun aandacht en motivatie tijdens observerend leren te versterken: voordat ze observeren hoe modellen een leestaak verrichten, moesten leerlingen eerst zelf deze leestaak verrichten. Daarmee hopen we hun engagement tijdens het observerend van de modellen te hebben versterkt.

Ten tweede hebben we in lijn met Braaksma (2002) de *evaluatieve* voorwaarde in onze lessenserie uitgebreid ruimte gegeven. Na het observeren van de modellen moesten leerlingen niet alleen het modelgedrag beschrijven, ze evalueerden bovendien dit modelgedrag op twee dimensies: 1) ze vergeleken de modellen zelf: welk model vertoonde een beter tekstbegrip, en wat kon het ‘zwakkere’ model van het ‘betere’ model leren?, en 2) ze vergeleken de modellen met hun eigen leesgedrag: in welk model herkenden ze zich, en wat konden ze zelf van de modellen leren? Op deze manier beoordeelden leerlingen in onze lessenserie expliciet hun eigen leesproces door hun gedrag met dat van de modellen te vergelijken, waarna ze hieraan verbeterpunten dienden te ontlenuen. Observerend-leren-studies die deze evaluatieve component in hun interventie meenemen, vinden vaak duidelijke effecten (vgl. Bouwer & Koster, 2015; Braaksma, 2002; Elving-Heida, 2019; Raedts et al., 2006; Rijlaarsdam et al., 2005, 2011); deze evaluatieve component ontbreekt daarentegen goeddeels bij observerend-leren-studies die nauwelijks of geen effect konden aantonen (vgl. Grenner et al., 2020; Van der Loo, 2022).

Het lijkt ons op basis van ons onderzoek goed als vervolgonderzoek verder verkent in hoeverre deze ‘versterkende’ componenten cruciale voorwaarden voor de effectiviteit van observerend leren zijn, én onder welke omstandigheden ze noodzakelijke toevoegingen hiervoor zijn.

Onderzoek naar de effectiviteit van soorten modellen. De studies in hoofdstuk 5 en 6 sluiten niet alleen aan bij het onderzoek naar de werkzaamheid van observerend leren, maar ook specifiek bij dat naar de werkzaamheid van *Eye-Movement Modeling Examples* in het onderwijs (zie Van Marlen, 2019). Het onderzoek naar de effectiviteit van EMME's binnen leesonderwijs is vooralsnog zeer beperkt (Salmerón & Llorens, 2019; Salmerón et al., 2020; vgl. Breukink, In voorbereiding), waarbinnen alleen Salmerón en Llorens (2019) vooralsnog een effect op het tekstbegrip van leerlingen konden vaststellen.

Recent werd ook de toepassingsmogelijkheid van *gaze-displays* voor het leesonderwijs nader verkend: deze statische weergaven van alle fixaties van een leerling op een tekst blijken docenten geen interpretatieproblemen op te leveren, waarmee ze mogelijk als bruikbaar didactisch instrument kunnen dienen (Knoop-Van Campen et al., 2021). Er lijkt echter een duidelijke inhoudelijke meerwaarde van EMME's boven dergelijke *gaze-displays*: in vergelijking daarmee staan EMME's immers leerlingen toe veel directer het leesproces van een model te volgen, met alle strategische stappen die dit model tijdens lezen zet – mits de EMME althans voorziet in duidelijk verhelderend audiocommentaar van het model (Jarodzka et al., 2013). Vervolgonderzoek naar specifiek de effectiviteit van EMME's voor het leesonderwijs lijkt al met al toe te juichen.

Het lijkt echter goed tegelijk te benadrukken dat niet alleen EMME's als een reflectie-instrument voor observerend leren kunnen gelden; ook andere instrumenten kunnen het denkproces van modellen veraanschouwelijken, zoals video-opnamen van hardop-denkende leeftijdsgenoten (bijv. Couzijn, 1995; Keehnen et al., 2015), of online-opnamen van (lees- of) schrijfhandelingen van modellen op

een beeldscherm, voorzien van hardop-denken (bijv. Van der Loo, 2022). Ook in ons onderzoek maakten we aanvullend gebruik van een ander instrument dan EMME's. In de derde observerend-leren-les in hoofdstuk 6 wilden we kunnen reflecteren op het taakmodel van twee modellen bij een complexe leestaak. Daarom gebruikten we hier niet een EMME of video – die vooral bruikbaar lijken om het schrijf- of leesproces zelf te veraanschouwelijken – maar stilstaande beelden van alinea's in een tekst die het model besprak, voorzien van diens hardop-denken – zodat we vooral de algemene aanpak van de leestaak door het model centraal konden stellen. Vervolgonderzoek naar welk instrument in welke situatie bij welk leerdomein het denkproces van modellen het best kan veraanschouwelijken, ligt dan voor de hand: op welke wijze kan bijvoorbeeld het gedrag van een leeftijdsgeenoot bij een uitgebreide complexe leestaak het best worden gevisualiseerd?

4 Implicaties voor het onderwijs

Wat betekent het onderzoek in dit boek voor het tekstbegripsonderwijs in het schoolvak Nederlands? In de discussieparagrafen van vooral de hoofdstukken 4 en 6 stonden we al vrij uitgebreid stil bij de implicaties voor het tekstbegripsonderwijs. Over het algemeen lijkt de stelling verdedigbaar dat het tekstbegripsonderwijs in het voortgezet onderwijs nog vaak aansluit bij wetenschappelijke inzichten van vóór de millenniumwisseling. In de inleiding stonden we al vrij uitgebreid stil bij drie voorbeelden hiervan. We bespraken daar:

- 1) de internationaal gezien nogal eigenzinnige invulling van strategie-instructie in leergangen Nederlands, een invulling die in grote lijnen afgeleid is van een Amerikaanse studeermethode uit 1946 (zie ook Rooijackers et al., 2021b) – hiermee staat het Nederlandse leesonderwijs internationaal op zichzelf;
- 2) de tekst-met-vragen-didactiek die vooral aansluit bij didactische inzichten uit de neerlandistiek uit de jaren zeventig en tachtig (bijv. Kreeft & Luijten, 1979): leerlingen zouden door begripsvragen te beantwoorden zelfstandig een tekst leren lezen – een aanname die niet op empirische bewijsvoering steunt;
- 3) de opvallend rigoureuze scheiding in veel leergangen tussen narratieve (en fictionele en literaire) teksten enerzijds en zakelijke teksten anderzijds, die eveneens haar oorsprong vindt in de jaren zeventig van de vorige eeuw (zie Griffioen, 1975) – een scheiding die theoretisch moeilijk verdedigbaar is (Kintsch, 1998).

In deze paragraaf zullen we uitgebreid stilstaan bij vier algemene implicaties voor het tekstbegripsonderwijs in het schoolvak Nederlands, die we elk zullen uitwerken in concrete aanbevelingen, gericht op met name docenten Nederlands uit het voortgezet onderwijs. Daarbij zullen we enerzijds recente theoretische inzichten bieden en anderzijds zo concreet mogelijke suggesties doen voor de lessen tekstbegrip in het voortgezet onderwijs, en dan met name voor vwo-leerlingen.

Mogelijk is deze concrete gerichtheid op de leesles Nederlands een stijlbreuk met voorgaande delen van dit boek; we hopen hiermee echter de bruikbaar-

heid van ons onderzoek voor vooral het klaslokaal Nederlands te optimaliseren. Sommige suggesties kunnen overigens zeker ook voor het onderwijs in moderne vreemde talen relevant zijn: veel van het hieronder genoemde ontbreekt bijvoorbeeld vrijwel in het recente handboek voor vreemdetalendidactiek (Dönszelmann et al., 2020).

We zullen in het vervolg ingaan op de volgende implicaties en aanbevelingen:

Implicatie 1: Maak het lezen van een tekst betekenisvol.

1. Zet de tekst met vragen als oefening terughoudend in.
2. Maak lezen betekenisvol en varieer met opdrachten.
3. Stel de diepe tekstbestudering voorop.

Implicatie 2: Reflecteer samen op het begripsproces.

4. Schenk meer aandacht aan het begripsproces van leerlingen zelf.
5. Maak van lezen een dialogische activiteit.
6. Reflecteer vanuit een functioneel analyse-instrumentarium.

Implicatie 3: Vergelijk, weeg en integreer informatie tussen tekst(del)en.

7. Schenk meer aandacht aan meervoudig-tekstbegrip.
8. Bied een breed scala aan (digitale) teksten aan.
9. Geef evalueren en reflecteren meer aandacht.

Implicatie 4: De rol van de docent in het leerproces is cruciaal.

10. Train docenten in procesgerichte klassikale reflectie.

4.1 Implicatie 1: Maak het lezen van een tekst betekenisvol

Als het om de begripsoperationalisatie van tekstbegrip gaat, sluit het huidige leesonderwijs in het schoolvak Nederlands goeddeels aan bij de grondslagen van de Constructie-Integratie-theorie (Kintsch, 1998). Dit model vertrekt vanuit het idee dat een vaardige lezer zorgvuldig één samenhangende, betekenisrijke tekst leest en zo een samenhangende mentale representatie ervan opbouwt. Deze theorie speelt zijdelings bijvoorbeeld een rol in de didactische handreiking van Hendrix en Hulshof (1994); onder meer op basis hiervan stellen zij een 'interactieve' benadering van teksten voor, waarin top-down-begripsprocessen worden afgewisseld met bottom-up-begripsprocessen. Ook een handboek als Bonset et al. (2019) sluit grotendeels aan bij deze begripsoperationalisatie. Huidige leergangen Nederlands bieden bovendien oefeningen aan die belangrijke uitgangspunten met de theorie delen: ze stellen vaak één samenhangende tekst uit de traditionele media centraal, geven geen betekenisvol leesdoel vooraf en zetten vooral in op precies lezen.

De Constructie-Integratie-theorie is nog altijd breed gedragen, maar ze kent duidelijke beperkingen: ze vertrekt vanuit een vrij ideale leessituatie (McNamara & Magliano, 2009a); ze voorspelt en verklaart daarnaast tekstbegrip vanuit een relatief eenvoudig, intratekstueel perspectief en stelt steeds het begrip van één tekst

centraal (Braasch et al., 2018); bovendien verdisconteert ze niet of nauwelijks de mogelijke impact van een leesdoel op het begrip van lezers (Linderholm & Van den Broek, 2002; Yeari et al., 2015). Soortgelijke kritiek kan voor veel huidige leer- gangen en handboeken voor het schoolvak Nederlands gelden: lezen is binnen de leesparagrafen van veel leergangen Nederlands doorgaans een min of meer doel- loze, zichzelf legitimerende handeling, zonder een realistische context ('Bestudeer de tekst en maak de vragen'). Dat is niet alleen een versmalling vanuit leestheo- retisch perspectief, het wreekt zich bovendien didactisch omdat in tal van andere onderwijssituaties lezen bij uitstek binnen betekenisvolle contexten gebeurt ('Be- studeer voor het proefwerk van volgende week het hoofdstuk over de Tachtigjarige Oorlog'). Bovendien gaat het voorbij aan het *first principle of instruction* dat leer- lingen levensechte, betekenisvolle taken moeten krijgen aangereikt (Merrill, 2002). Een eerste implicatie van ons onderzoek is daarom: 'Maak het lezen van een tekst betekenisvol'. We werken deze implicatie uit in drie aanbevelingen.

Aanbeveling 1: Zet de tekst met vragen als oefening terughoudend in

De traditionele tekst-met-vragen-didactiek in de leesles Nederlands heeft in ons onderzoek veel aandacht gekregen. De hoofdstukken 2, 3 en 4 leverden onderbou- wing op voor een aantal oude kritiekpunten op deze dominante taak in de leesles (bijv. Hoogeveen & Bonset, 1998). We lopen vijf van deze kritiekpunten en de on- derbouwing ervoor vanuit ons onderzoek langs.

- *Leerlingen leren "niet een tekst goed te analyseren"* (Van de Wetering & Groe- nendijk, 2015, p. 6). Die kritiek lijkt op basis van ons onderzoek aannemelijk: leerlingen lezen in hoofdstuk 2 en 3 de teksten vooraf nogal lineair en opper- vlakkig door. Bovendien suggereren de resultaten in hoofdstuk 4 dat een leer- ling die een tekst vooraf intensief bestudeert niet vaker een goed antwoord op de vragen zal hebben, dan een leerling die dat niet of minder intensief doet.
- *"Dat is wat we ze leren: trucjes"* (Heemskerk, 2013). In hoofdstuk 3 gaf een deel van de leerlingen ongevraagd aan dat ze een antwoord op een begripsvraag geheel baseerden op conventies rondom de tekst met vragen ('De eerste vraag gaat altijd over de eerste alinea'). 'Trucjes' lijken daarmee inderdaad bij deze taak op de loer te liggen.
- *"Je meet slechts hoe handig leerlingen zijn in het geven van gewenste antwoor- den"* (Welgraven & De Vos, 2019). In hoofdstuk 2 en 3 bleken leerlingen zeer veel energie in het antwoordproces te stoppen, maar niet zozeer in het voor- af-leesproces. In hoofdstuk 4 zagen we dat de tekstraadpleegbaarheid bij de vragen een groot effect heeft op de antwoordscore, terwijl een intensieve vooraf-tekstbestudering juist géén aantoonbaar effect heeft. We concludeer- den daar dat het bij deze taak draait om het goed kunnen beantwoorden van de vragen; het vooraf-lezen is vooral een instrument waarmee leerlingen ge- makkelijker de vragen kunnen interpreteren en antwoord-relevante tekst- delen kunnen lokaliseren. De activiteiten tijdens de antwoordfase geven dus hoogstwaarschijnlijk de doorslag voor de inhoud van het antwoord.
- *"Het is onwaarschijnlijk dat leerlingen op deze wijze zullen leren om zelfstan-*

dig om te gaan met informatie" (Hoogeveen & Bonset, 1998, p. 232). Op basis van het oppervlakkige leesgedrag dat we in hoofdstuk 2 en 3 beschreven, stelden we daar al vast dat deze kritiek aannemelijk is. In hoofdstuk 4 concludeerden we dat een zelfstandige, diepgaande vooraf-bestudering bij de reguliere begripsvragentoets (naar alle waarschijnlijkheid) niet rendeert voor de antwoordscore; wie de moeite neemt de tekst goed te bestuderen, ziet dat bij deze toets naar alle waarschijnlijkheid niet in een hogere antwoordscore vertaald. Als leerlingen in de les vaak oefenen met deze toetsvorm, ligt het dus niet voor de hand dat ze daarmee zelfstandige goede lezers worden.

- En, samenvattend: *het huidige leesonderwijs is "toetsgericht en niet ontwikkelingsgericht"* (Hoogeveen & Bonset, 1998, p. 232). De veelvuldige inzet van deze taak lijkt inderdaad sterk op *teaching to the test*: leerlingen worden ertoe aangezet het concept 'tekstbegrip' te reduceren tot het kunnen geven van een goed antwoord op een begripsvraag - terwijl deze taak weinig ecologische validiteit kent: hoe vaak moeten immers volwassen Nederlanders een serie begripsvragen beantwoorden bij een tekst? Dat docenten zich graag richten op het meetbaar controleren van leerlingantwoorden, is niet vreemd: de behoefte aan meetbaarheid in de tekstbegripsles is immers een oude, bekende valkuil (Durkin, 1978). Tekstbegrip is echter veel meer dan een eenduidig te beoordelen antwoord op een tekstafhankelijke begripsvraag. Zoals we in hoofdstuk 3 al concludeerden: er lijkt hier al met al sprake van *survival of content that fittest best to the high-stakes test* (Popham, 2001). Oftewel: toetsing en training worden hier met elkaar verward. Onze conclusie luidt daarom: wees terughoudend met deze taak. Je oefent ermee uiteindelijk vooral het beantwoorden van vragen bij een tekst - maar waarschijnlijk geen zelfstandig, diep, geconcentreerd lezen.

Ons onderzoek geldt overigens in strikte zin enkel vwo-leerlingen; er lijkt ons echter geen reden waarom deze zelfde terughoudendheid niet voor alle afdelingen in het voortgezet onderwijs kan gelden. Het is immers maar zeer de vraag of gemiddeld minder vaardige lezers zich 'wenselijker' zullen gedragen dan vwo-leerlingen. (Gezien geluiden van docenten Nederlands dat bijvoorbeeld vmbo-leerlingen het vooraf-lezen bij teksten met vragen vaak overslaan, lijkt het tegendeel zelfs eerder hier te verwachten.)

Aanbeveling 2: Maak lezen betekenisvol en varieer met opdrachten

We bespraken het al hierboven: vervang het context- en doelloos lezen van een tekst door een functionele, levensechte context, waardoor het lezen van een tekst voor leerlingen betekenisvol en vanzelfsprekend wordt, het lezen een 'tastbaar' doel krijgt en de set begripsvragen niet langer een lukrake 'grabbelton' lijkt. Dat kan op tal van manieren.

Het kan allereerst door de tekst-met-vragen-taak van een expliciet, betekenisvol doel of een betekenisvolle context te voorzien en de set begripsvragen hierop logisch te laten aansluiten. Zo kunnen docenten vooraf expliciet een passend realistisch doel bij een tekst formuleren ('Je wilt deze betogende tekst over orgaandonatie gebruiken als uitgangspunt voor een debat over verplichte orgaandonatie voor volwassenen') en

alleen begrips-, evaluatie- en reflectievragen stellen die logischerwijs volgen vanuit dit doel. Het ligt voor de hand hierbij regelmatig teksten te kiezen waarmee leerlingen in het leven van alledag als vanzelf te maken kunnen hebben. Men denke aan een tekst van een overheidssite waarin de wettelijk vastgelegde arbeidsvoorwaarden en -omstandigheden voor jongeren onder de achttien jaar worden beschreven, en aan begripsvragen die ze ertoe aanzetten alle voor hen relevante informatie te identificeren.

Een andere logische mogelijkheid is het integreren van leestaken in andere taal-domeinen, zoals schrijven of spreken. Een pittige, interessante tekst kan bijvoorbeeld als uitgangspunt dienen voor een gezamenlijke, klassikale schrijfopdracht over een onderwerp uit de taalbeschouwing, Nederlandse maatschappij, cultuur of communicatie, waarbij leerlingen ernst, gevolgen, problemen en oplossingen rond die problematiek nauwkeurig in kaart brengen. Dit kan natuurlijk ook gekoppeld worden aan leerinhoud binnen het schoolvak Nederlands: zijn leerlingen in staat alle verklaringen voor het uitsterven van dialecten in Nederlands in een leertekst aan te wijzen en in eigen woorden weer te geven?

Aanbeveling 3: Stel de diepe tekstbestudering voorop

Nederlandse leerlingen zijn relatief goed in informatie zoeken (Gubbels et al., 2019); ze moeten echter ook diep, geconcentreerd kunnen lezen, gericht op de samenhang en betekenis van een tekst (Onderwijsraad & Raad voor Cultuur, 2019; vgl. Vernooy, 2022). Diep lezen zorgt ervoor dat leerlingen een actieve mentale representatie van de tekst construeren, zodat ze daarna informatie kunnen interpreteren, gevolgen eraan kunnen verbinden, kunnen reflecteren op de inhoud en geloofwaardigheid ervan en informatie kunnen evalueren en toepassen op andere situaties (Wolf, 2018).

Leerlingen moeten daarom in de leesles vaak geconcentreerd, herhaaldelijk en langdurig een pittige tekst doorgronden, individueel, in groepjes of klassikaal. Daarbij ligt het voor de hand te werken vanuit een helder leesdoel, al dan niet binnen een grotere, domeinoverstijgende taak zoals een debat of schrijfopdracht. Leerlingen kunnen bijvoorbeeld eerst informatie uit een tekst schematisch in kaart brengen via een *mind map*, waarna ze samen over hun bevindingen praten, en/of klassikaal de voorwaarden bespreken waaraan de *mind map* moet voldoen, en de manier bespreken waarop ze deze voorwaarden in de tekst hadden kunnen vinden (vgl. Robben, 2012). Daarbij past dan een uitdagende vervolgtask, volgend vanuit het vooraf-gegeven leesdoel: leerlingen kunnen bijvoorbeeld in een schrijftask kritisch op de onderbouwing in de tekst reflecteren of de gedachtegang van de tekst toepassen op een andere, niet in de tekst beschreven situatie.

Cruciaal is dat leerlingen het eindproduct alleen via een grondige, diepgaande tekstverwerking afdoende kunnen afronden. Het stellen van hoge, duidelijke eisen ('Deze tekst bevat acht verklaringen; wijs ze *allemaal* aan door ze in de tekst te onderstrepen, en geef ze in eigen woorden daarna zo nauwkeurig mogelijk weer in je schrift') lijkt ons daarbij een belangrijk didactisch instrument.

Een ander instrument, waarop wij in ons onderzoek regelmatig zijn teruggevalen, noemen we hier graag nog met name. Regelmatig hebben we in ons onderzoek leerlingen eerst een tekst vanuit een specifiek doel laten lezen, waarna we de

tekst innamen en een – op het doel aansluitende – verwerkingsopdracht aanboden. Het kan een krachtig instrument vormen om leerlingen zelfstandig teksten te laten bestuderen; hun prestatie op de verwerkingstaak vormt voor leerlingen vervolgens een helder controlemiddel voor de kwaliteit van hun tekstbestudering (vgl. Johnston, 1984; Schroeder, 2011).

4.2 *Implicatie 2: Reflecteer samen op het begripsproces*

De kracht van de *social learning theory* van Bandura is dat het voor leerprocessen de meerwaarde erkent van de observatie van anderen; bij die observatie draait het dan niet zozeer om het uiteindelijke product als wel om het rationale erachter, om de processen die eraan ten grondslag liggen (Bandura, 1986; Bandura & Walters, 1977). Observerend leren is daarmee een krachtig instrument gebleken om (reflectie op) complexe leerprocessen te didactiseren.

Over het algemeen geldt dat de aanpak in de huidige leesles Nederlands zich nog vaak vooral richt op het product dat een leerling na het lezen van een tekst maakt: de antwoorden op begripsvragen. Zijdelings krijgt in de les *modeling* aandacht: docenten doen dan voor hoe ze zelf als experts een tekst lezen (Bogaerds-Hazenbergh & Evers-Vermeul, 2020; Linthorst & De Glopper, 2015). Op *peer*-niveau lijkt er echter weinig op teksten te worden gereflecteerd. Dat valt op omdat zo'n meer procesgerichte benadering vanuit leerlingen wel ingang heeft gevonden in het domein schrijfvaardigheid in het schoolvak Nederlands: in leergangen Nederlands worden leerlingen doorgaans uitgenodigd elkaars schrijfproducten te beoordelen en met feedbackformulieren van commentaar te voorzien, waarna ze hun eigen product moeten herschrijven (bijv. Frank et al., 2019; Merckx et al., 2018). De laatste jaren is er ook veel aandacht voor de kracht van deze comparatieve vergelijking bij schrijfvaardigheidsonderwijs, waarvoor digitale leeromgevingen inmiddels rijke toepassingsmogelijkheden bieden (vgl. Bouwer et al., 2018a).

Voor het leesonderwijs lijkt hier nog een wereld te winnen. Immers, niet alleen schrijftaken, ook leestaken kunnen leerlingen voor complexe handelingsproblemen en zelfregulerende beslissingen plaatsen. Het is daarom goed als docenten en leerlingen samen in de leesles meer over het begripsproces praten, en de relatie die dat begripsproces heeft met het eindproduct. We werken deze implicatie weer uit in drie aanbevelingen.

Aanbeveling 4: Schenk meer aandacht aan het begripsproces van leerlingen zelf

Zoals we hierboven al stelden, wordt er op dit moment in het Nederlandse leesonderwijs relatief beperkt aandacht besteed aan het begripsproces van leerlingen. Alleen *modeling* door docenten heeft een zekere ingang gevonden. Echter, het is goed om leerlingen naast deze *mastery models* vooral ook *coping models* aan te bieden; de laatste blijken in de les vaak effectiever dan de eerste, omdat leerlingen de problemen en oplossingen van deze leeftijdsgenoten herkennen en dat voorbeeldgedrag daarmee makkelijk op zichzelf betrekken (Schunk, 1987; Zimmerman & Kitsantas, 2002; Schunk & Zimmerman, 2007).

In het Nederlandse leesonderwijs zou er dus meer aandacht mogen zijn voor de manier waarop leerlingen teksten lezen. Docenten zouden regelmatig met leerlingen over hun leesprocessen moeten praten en over de strategieën die ze daarbij inzetten. We merken daarbij drie aandachtspunten op.

Ten eerste: de generieke ‘manieren van lezen’ die in leergangen en handboeken Nederlands nog vaak strategie heten (‘oriënterend lezen’, ‘globaal lezen’, etc.), zouden hierbij naar ons idee geen rol moeten spelen: deze hebben immers geen betrekking op strategieën, op de middelen die een lezer actief inzet om een tekst te begrijpen, maar op generieke manieren van lezen. In de inleiding stelden we in § 1.3 vast dat Nederlandse leergangen met deze eigenzinnige vorm van strategie-instructie vrijwel op zichzelf staan. Docenten en leerlingen zouden daarentegen moeten praten over evidencebased leesstrategieën, over de bewuste, doelgerichte middelen die lezers inzetten, zoals het herlezen van kernzinnen, het tussentijds samenvatten van de tekstuele gedachtegang en het tussentijds voorspellen van de tekstinhoud – een aanpak zoals die in veel internationale strategie-instructie is te vinden (vgl. Almasi & Fullerton, 2012; Duke et al., 2011; Moeken et al., 2015; Rooijackers et al., 2021b).

Ten tweede: het is goed op te merken dat leerlingen – in onze ervaring – niet altijd even gemakkelijk hun eigen leesproces kunnen beschrijven en hierop kunnen reflecteren, waardoor het moeilijk kan zijn hierover een goed klassengesprek te voeren. Animaties, video’s, EMME’s of andere visualisatievormen kunnen dan een hulpmiddel vormen om hierover toch een klassengesprek te voeren, zoals de EMME’s in de lessenserie die we beschreven in hoofdstuk 5 en 6.

Ten derde: een klassengesprek over het lezen van leerlingen hoeft niet alleen op hun begripsprocessen betrekking te hebben; het is ook goed met leerlingen te reflecteren op hun context- en taakmodellen bij complexe leestaken. Deze kunnen lezers immers eveneens zeer variabel invullen (Britt et al., 2018). Indien leerlingen argumenten moeten verzamelen in een set artikelen (voor een te schrijven betoog of te houden debat), waaraan moet hun eindproduct dan voldoen? Hoe selecteren ze de meest bruikbare artikelen? Hoe selecteren ze in de bruikbare artikelen de relevante argumenten? En hoe brengen ze alle verzamelde argumenten handig in kaart? Welke stappen dienen ze tijdens dit gehele proces dan idealiter te hebben doorlopen? Dergelijke complexe denkprocessen maken immers evenzeer deel uit van tekstbegrip in een bredere zin.

Aanbeveling 5: Maak van lezen een dialogische activiteit

Er is in het huidige leesonderwijs in het voortgezet onderwijs een nogal exclusieve rol weggelegd voor individueel stil-lezen: een leerling moet vaak zelfstandig zich een tekst eigen maken, waarna een leestaak volgt. Zoals we in § 1.1 van de inleiding al vaststelden, contrasteert de praktijk van nu daarmee drastisch met die in de periode 1900-1940 toen er in het klaslokaal Nederlands veel en vaak hardop gezamenlijk werd gelezen; stil-lezen kreeg in Nederland pas in het klaslokaal vanaf de jaren dertig een prominente plek (vgl. Diels, 1938; Langeveld, 1934).

Hardop lezen is echter belangrijk omdat het interactie toestaat over het lezen tussen docent en leerlingen onderling. Die interactie is een cruciale schakel in ef-

fectief leesonderwijs (Fisher & Frey, 2015) en in bredere zin in een goede taal-denkontwikkeling van leerlingen (Damhuis & De Blauw, 2008). Via discussie en sociale interactie over (het begrijpen van) teksten leren jongeren hoe ze meer diepgang in hun tekstbegrip kunnen brengen en hoe ze een complexe leestaak goed kunnen afronden. Wilkerson en Son merken in dit opzicht op dat docenten zich in de leesles niet moeten richten “on increasing the amount of student talk per se, but on enhancing the quality of the talk” (2011, p. 371). Door samen in een klas hardop te lezen en na te denken over wat er wordt gelezen, ontstaat bij leerlingen beter tekstbegrip.

Er lijkt dus veel voor te zeggen in de lessen tekstbegrip vaker dan nu hardop te lezen op te nemen; binnen de aanstaande vernieuwing van het schoolvak Nederlands komt hiervoor waarschijnlijk meer ruimte (SLO, 2021a, 2021b).

Aanbeveling 6: Reflecteer vanuit een functioneel analyse-instrumentarium

Het hoeft in de lessen tekstbegrip natuurlijk niet steeds over ‘zuiver’ begrip te gaan, er is veel voor te zeggen leerlingen voor een beter tekstbegrip een analyse-instrumentarium te leren gebruiken. Het is dan bijvoorbeeld belangrijk dat leerlingen de tekststructuur leren herkennen; onderwijs in algemene tekstuele basisstructuren, in alineaopbouw en in signaalwoorden vormt een effectief onderdeel van tekstbegripsonderwijs (bijv. Bogaerds-Hazenberg et al., 2022a; Van den Branden et al., 2019; Houtveen et al., 2019).

Momenteel vormen vragen over tekststructuur in lees oefeningen in leergangen Nederlands nog vaak een ‘toevallig’ onderdeel in een ogenschijnlijk lukraak samengestelde set vragen over allerlei tekstdimensies (argumentatieschema’s, drogredenen, woordenschat, inferenties, etc.). De vraag is in hoeverre deze bevragingwijze leerlingen de praktische meerwaarde van een dergelijk functioneel-analytisch instrumentarium laat inzien en of het onderwijs in de toepassing van analytische begrippen voor leerlingen niet vaak een doel op zichzelf blijft. Het viel ons tijdens het afnemen van het experiment vooral in hoofdstuk 5 op hoezeer vwo-leerlingen over het algemeen heel redelijk in staat bleken bij de leestaken in de werklessen de belangrijkste inhoudelijke uitspraken in een tekst te isoleren en onder woorden te brengen; dezelfde bevinding rapporteerden deelnemende docenten tijdens het experiment in hoofdstuk 6. Echter, docenten van drie scholen gaven in dit experiment ook aan dat leerlingen voorafgaande aan de eindtoets graag leerstof wilden: konden we leerlingen geen lijst functionele begrippen aanbieden, zodat ze “wat te leren hadden” en de leerstof “moeilijk” werd?

De vraag is of zo’n op zichzelf staande didactiek van tekststructurering wenselijk is; het laat immers het sterk contextuele (conditionele) karakter van tekststructurering buiten beschouwing (bijv. Bogaerds-Hazenberg et al., 2022a). Het lijkt zinniger tijdens het leesproces, het gezamenlijk lezen van de tekst, het herkennen van structurerende elementen in een tekst te bespreken: welke specifieke strategieën zetten ze bij deze specifieke tekst in om structurerings-elementen te herkennen? En hoe kunnen ze deze structurering voor zichzelf overzichtelijk verwerken? Moeten ze bijvoorbeeld pijlen in de kantlijn noteren om causale verbanden tussen alinea’s in kaart te brengen? Of moeten ze nummers vóór alinea’s zetten zodat alle argu-

menten voor een stelling in de tekst makkelijk terug te vinden zijn? En waarom is het voor deze tekst en voor het leesdoel belangrijk dergelijke structurerende strategieën in te zetten? Dat geldt eveneens voor eventuele verwerkingsopdrachten die docenten ná het lezen aanbieden: het kunnen herkennen en aanwijzen van tekst-structurerende elementen dient naar onze smaak logisch ingebed te zijn binnen leesdoel en opdracht, behalve als het analyse-instrument zelf onderwerp van de leesles is.

4.3 *Implicatie 3: Vergelijk, weeg en integreer informatie tussen tekst(del)en*

Sinds de millenniumwisseling zijn leesonderzoekers internationaal steeds meer het klassieke paradigma van ‘vaardige lezer bestudeert gezaghebbende tekst’ à la Kintsch (1998) gaan uitbreiden tot een paradigma dat wisselende leesdoelen, wisselende soorten lezers, wisselende soorten teksten en documenten verdisconteert (Fox & Alexander, 2017). Het internationale onderzoek naar meervoudig-tekstbegrip (*multiple text comprehension*) springt daarbij in het oog. Binnen de theorievorming rondom meervoudig-tekstbegrip gaat men ervan uit dat een tekst niet alleen een serie talige beweringen is à la Kintsch (1998), maar vooral ook een sociale entiteit: een boodschap, gekenmerkt door situationele, retorische en communicatief-pragmatische dimensies (Britt & Rouet, 2012; Britt et al., 2018; Rouet & Britt, 2011).

De implicaties van dit onderzoek zijn zeer relevant voor Nederlandse leerlingen. In vooral hoofdstuk 3 en 4 stelden we al de beperkingen vast van de huidige focus in het schoolvak Nederlands op enkelvoudig-tekstbegrip. Nederlandse vijftienjarige leerlingen hebben hoogstwaarschijnlijk moeite met het reflecteren op teksten, het integreren van informatie uit meerdere teksten en het evalueren van de betrouwbaarheid en overtuigingskracht van een tekst (Van den Broek et al., 2021; Gubbels et al., 2019). Er weerklinken in Nederland dan ook al geruime tijd geluiden deze meer 21e-eeuwse dimensies een prominentere plek te geven in de lessen tekstbegrip in het middelbaar onderwijs (bijv. Van den Broek et al., 2021; Gubbels, 2020; Kirschner, 2017; Rooijackers et al. 2021a; SLO, 2021b). In de Centraal Examens Nederlands havo/vwo zullen deze dimensies vanaf 2024 waarschijnlijk een plaats krijgen (Evers-Vermeul et al., 2022). Op dit moment hebben recente Nederlandse handboeken en Nederlandstalige overzichtsstudies voor het leesonderwijs nog relatief weinig oog voor meervoudig-tekstbegrip, noch voor de specifieke kennis en vaardigheden die digitale tekstsoorten van lezers vergen (vgl. Bonset et al. 2019; Van den Branden et al., 2019; Houtveen et al., 2019; Ros et al., 2021). Bij de overzichtsstudies komt dat waarschijnlijk doordat ze zich (al dan niet impliciet) vooral op het basisonderwijs richten. Daarin wordt het begrijpen van meerdere teksten of (digitale) bronnen over één onderwerp mogelijk nog weinig relevant geacht – een aanname die evenwel ter discussie is gesteld (Van den Broek et al., 2021).

Ook hier lijkt voor de leeslessen Nederlands nog een wereld te winnen. Er mag meer gewerkt worden met in kwaliteit en vorm wisselende teksten en documenten over één onderwerp. Daarbij is het vergelijken van (digitale) informatie, het wegen van informatie én synthetiseren van (digitale) informatie bij het werken met een of

meerdere teksten een belangrijk aandachtspunt. We werken deze implicatie weer uit in drie aanbevelingen.

Aanbeveling 7: Schenk meer aandacht aan meervoudig-tekstbegrip

Meervoudig-tekstbegrip is een complexe, gelaagde vorm van begrip. Perfetti et al. (1999) en Britt en Rouet (2012) stellen dat lezers bij het lezen van meerdere documentsoorten over één onderwerp (teksten, grafieken, podcasts, video's etc.) eerst per bron afzonderlijk een samengestelde mentale representatie van de tekstinhoud, een 'situatiesmodel' (*Situations Model*), in hun hoofd opbouwen. Gelijktijdig construeren ze een intertekstmodel (*Intertext Model*), een model dat per document de meer sociale dimensies achter beweringen meeneemt: achtergrond, gezag en motieven van de auteur, herkomst en leeftijd van de publicatie, genre, aangesproken publiek. Uiteindelijk kunnen lezers zo over het onderwerp een geïntegreerd, passend mentaal model (*Integrated Documents Model*) opbouwen, één logische, op bruikbaarheid gewogen, samenhangende representatie van de informatie van alle documenten samen. Het RESOLV-model van Britt et al. (2018) vult deze ideeën nog verder aan, door hierin ook de taakcontext te verdisconteren. Bij een leestaak vertrekken lezers, aldus dit model, vanuit een statisch contextmodel, dat zaken verdisconteert als de taakinterpretatie, de status van de 'opdrachtgever', de hulpmiddelen en belemmeringen bij de taak, evenals een zelfbeoordeling bij deze taak door de lezers. Vanuit dit contextmodel stellen lezers vervolgens een dynamisch taakmodel op: taakspecifieke doelen en scenario's.

Leerlingen zouden vaker dan nu het geval lijkt, moeten oefenen met taken waarbij ze meerdere documenten over één onderwerp moeten verwerken. Leerlingen zouden daarbij moeten oefenen met het opbouwen van een passend geïntegreerd mentaal model bij deze teksten, waarbij ze niet alleen op de inhoud en argumentatieve aanvaardbaarheid van één tekst reflecteren, maar ook diepgaand de inhoud en kwaliteit van bronnen vergelijken en evalueren.

Docenten moeten bij dit soort taken aandacht geven op verschillende niveaus. Het begint bij het opstellen van passende *context- en taakmodellen* door leerlingen: wat vraagt een taak van leerlingen? Aan welke voorwaarden zal hun eindproduct moeten voldoen? Welke stappen zullen ze idealiter daarvoor moeten inzetten? Hoe bouwen ze tijdens het verrichten van de taak controlemomenten in, zodat ze op de tussenstand kunnen reflecteren en het proces tot nu toe kunnen evalueren?

Vervolgens moeten docenten aandacht geven aan het opbouwen van *situaties- en intertekstmodellen*. Hoe maken leerlingen de informatie in de documenten tot een logisch, samenhangend geheel? Hoe scheiden ze kaf van koren, en welke criteria gebruiken ze daarbij? Dit soort taken zijn voor veel lezers ronduit lastig. Omdat minder vaardige lezers zo'n geïntegreerd mentaal model opstellen, blijken ze er namelijk goed aan te doen eerst taken te verrichten waarmee ze de inhoud van elke tekst afzonderlijk verwerken, zoals samenvattingen of schema's (Gil et al., 2010). Maar ook geoefende lezers blijken vaak behoefte te hebben aan begeleiding: ze doen er goed aan aantekeningen en schema's te maken om informatie overzichtelijk te ordenen (Britt & Rouet, 2012).

Aanbeveling 8: Bied een breed scala aan (digitale) teksten aan

Leerlingen van vandaag de dag moeten zich kritisch en zelfstandig een weg kunnen banen in de 21e-eeuwse online informatiestroom. Er is wel gesteld dat informatiegeletterdheid bij uitstek dé 21e-eeuwse vaardigheid is (Kirschner, 2017). Echter, in veel huidige leesparagrafen van leergangen Nederlands is er vooralsnog slechts geringe aandacht voor ‘nieuwe’ soorten digitale bronnen; ‘klassieke’ (lineaire, unimodale, continue) teksten uit traditionele papieren media overheersen momenteel in veel leergangen Nederlands en (vooralsnog) in de (centrale) toetsing van tekstbegrip in het schoolvak Nederlands – al staat dit te veranderen (Evers-Vermeul et al., 2022). In de lessen Nederlands moet er echter ruim aandacht zijn voor deze digitale internetbronnen en conventies, opbouw, meerwaarde en beperkingen ervan.

Docenten doen er goed aan in de klas leerlingen te confronteren met een breed palet aan tekstsoorten, zoals die online en offline op hun pad kunnen komen. Het is belangrijk in de les het onderscheid te bespreken tussen het lezen van multimodale of niet-lineaire teksten (internetsites) en dat van lineaire teksten (krantenartikelen), waarbij leerlingen oefeningen krijgen waarmee ze met deze tekstsoorten aan de gang kunnen. Het is bovendien goed in de les regelmatig kwalitatief sterk uiteenlopende bronnen aan te bieden; naast artikelen uit traditionele gezaghebbende media kunnen dan bijvoorbeeld posts op een webforum over hetzelfde onderwerp prijken. In de les moeten dan vragen voorbijkomen als: hoe vind je bruikbare, betrouwbare, relevante informatie op internet en hoe scheid je relevante van irrelevante informatie? In een wereld waarin nog weleens een ‘infodemie’ lijkt te heersen, vol mis- en desinformatie (WHO, 2022), is het cruciaal dat leerlingen leren reflecteren op de kwaliteit van bronnen – waarmee de volgende aanbeveling zich als vanzelf aandient:

Aanbeveling 9: Geef evalueren en reflecteren meer aandacht

In de twintigste eeuw werden leerlingen vaak vanzelfsprekend geconfronteerd met papieren teksten uit gezaghebbende media; in de 21e eeuw krijgen leerlingen via internet en sociale media een uiterst brede waaier aan teksten en documenten onder ogen, sterk wisselend in kwaliteit (bijv. Kirschner, 2017), waarbij hun waardering voor traditionele papieren media anno 2021 niet erg hoog lijkt (Dera & Van Doeseelaar, 2022). Het belang van het kunnen vaststellen van de waarheid, betrouwbaarheid, bruikbaarheid en relevantie van informatie van bronnen wordt internationaal breed onderschreven: het heeft het laatste decennium bijvoorbeeld een prominente plaats gekregen in internationale leestoetsen zoals Pirls en PISA (vgl. Mullis et al., 2017; OECD, 2019): “Evaluation and reflection were arguably always part of reading literacy, but their importance has increased with the increased amount and heterogeneity of information readers are faced with today” (OECD, 2019, p. 36). Binnen het subdomein ‘evalueren en reflecteren’ moeten leerlingen voor de PISA-toets daarom niet alleen kunnen “reflecteren op inhoud en vorm” van teksten, maar ook “de waarheid en geloofwaardigheid” van teksten kunnen vaststellen, en ze moeten binnen meerdere teksten over één onderwerp “tegenstrijdigheden” kunnen signaleren en deze in hun begrip een plaats kunnen geven (OECD, 2019).

Er lijkt alle reden om deze subdimensie van tekstbegrip een prominentere plaats te geven in de leeslessen Nederlands dan nu het geval is. We zeiden het al enkele malen in dit boek: vooral op de subdimensie ‘evalueren en reflecteren’ scoren Nederlandse leerlingen ondermaats.

Leerlingen moeten dus allereerst nadenken over de *betrouwbaarheid* van informatie. Docenten kunnen leerlingen bijvoorbeeld een serie bronnen over één onderwerp aanbieden – met of zonder (!) bronvermelding –, waarin ze een rangorde moeten aanbrengen in betrouwbaarheid en hun criteria moeten expliciteren: zo’n opdracht laat leerlingen stilstaan bij inhoudelijke kenmerken (taalgebruik, vereiste voorkennis, scherpthe in redeneringen, brongebruik in de bron, etc.) en bij bronkenmerken (gezag, partijdigheid van de schrijver, vindplaats van de bron, etc.). Leerlingen kunnen nadenken over *bruikbaarheid en relevantie* van informatie. Zo kunnen docenten bijvoorbeeld een internettekst mét alle gebruikersreacties aanbieden: welke reacties zijn relevant en waarom? Hoe kunnen ze herkennen welke reacties relevant zijn en welke niet? Of leerlingen ontvangen vanuit een duidelijk leesdoel (bijvoorbeeld een debat over een maatschappelijk onderwerp) een diverse set bronnen over één onderwerp, waarbij ze moeten aangeven welke bronnen bruikbaar zijn en welke niet. Daarbij moeten ze *tegenstrijdigheden* tussen deze bronnen kunnen herkennen en (idealiter) deze kunnen verklaren. En ten slotte: leerlingen zouden opdrachten moeten ontvangen waardoor ze de gedachtegang van een tekst *op een hoger denkniveau moeten toepassen*. Zo kunnen ze opdrachten maken waarbij ze gedachtegang in de tekst moeten toepassen op een verwante, niet expliciet in de tekst benoemde situatie: wat betekenen de uitspraken over situatie A in tekst X voor de verwante situatie B? Of ze kunnen bij een probleemoplossingstekst stilstaan bij de mogelijke bezwaren tegen de geopperde oplossingen, op grond van hun eigen kennis over het onderwerp.

Door te werken met een breed scala aan teksten en documenten kunnen docenten leerlingen op een krachtige manier laten nadenken over de waarde van informatie, zodat leerlingen kunnen leren zich als zelfdenkende burgers overeind te houden in de moderne informatiestortvloed.

4.4 Implicatie 4: De rol van de docent in het leerproces is cruciaal

In hoofdstuk 6 zagen we dat de docent een cruciale meerwaarde heeft voor het leerproces van leerlingen: leerlingen die zelfstandig de lessenserie doorwerkten, steken gemiddeld minder vaak wat op van de lessenserie dan leerlingen die de lessenserie onder leiding van de docent doornamen. Daar stelden we vast dat tijdens de Covid19-epidemie de cruciale rol van de docent in het leerproces nog weleens buiten beschouwing werd gelaten doordat door de schoolsluitingen thuiswerken voor leerlingen de norm was (Smeets, 2021). Onze studie liet zien dat in de leersituatie zonder docent leerlingkenmerken bepalen of ze tot leren komen.

Dat lijkt een belangrijke les. Buiten de Covid19-epidemie om zijn docenten Nederlands de laatste tien jaar steeds meer gaan beschikken over digitaal lesmateriaal: heel wat leergangen Nederlands bieden inmiddels een omvangrijke digitale

leeromgeving waarin leerlingen zelfstandig verdiepende of versterkende modules over een bepaald vakonderdeel kunnen maken. De vraag is of overdaad hier niet al snel schaadt. In het overzichtswerk van Hattie (2012) wordt omstandig aangetoond dat de docent in de klas – door aanwezig te zijn, te handelen, te informeren en instrueren, in werkvormen te wisselen, (klassen)gesprekken te voeren, etc. – altijd bijdraagt aan het leren van leerlingen.

Dat geldt binnen ons onderzoek uit hoofdstuk 6 specifiek voor het reflecteren en evalueren op het eigen leesgedrag door leerlingen. Als docenten niet de klas-sengesprekken leiden, als ze niet een gezamenlijke klassikale reflectie en evaluatie regisseren en stimuleren, is het maar de vraag of leerlingen gemiddeld iets van observerend leren opsteken. Met name in een gezamenlijke reflectie en evaluatie kan observerend leren een waardevol didactisch instrument zijn.

Wel is het voor ons de vraag in hoeverre docenten Nederlands op dit moment voldoende beslagen ten ijs komen, willen ze klassikaal effectief het begripsproces van leerlingen bij leestaken analyseren, erop reflecteren en het evalueren. In de les-observaties in hoofdstuk 6 viel het op dat docenten het niet eenvoudig vonden in detail met een klas op het leesgedrag van de modellen in de filmpjes te reflecteren. De reflectie was soms nogal basaal en ging maar beperkt in op details in het gedrag van de modellen; ook in de klassikale evaluatie deden docenten niet altijd recht aan de modellen. Zo bevestigde een docent leerlingopmerkingen dat een model inderdaad té snel las om de tekst goed te kunnen begrijpen, terwijl uit het hoorbare modelcommentaar bleek dat deze de tekst treffend in eigen woorden kon samenvatten. Docenten hadden ook regelmatig moeite scherp de strategieën aan te wijzen en onder woorden te brengen waarvan de modellen gebruikmaakten. Meer dan bij de voorgaande aanbevelingen, lijkt hier ons een zekere mate van (na)scholing noodzakelijk. Daarom sluiten we af met de aanbeveling het handelingsreper-toire van de docent Nederlands specifiek hierop uit te breiden:

Aanbeveling 10: Train docenten in procesgerichte klassikale reflectie

Zeker in het domein leesvaardigheid zijn leerlingen in het voortgezet onderwijs vaak nogal productgericht. In het basisonderwijs hebben veel leerlingen vaak tekst-begrip geoefend met teksten met vragen (bijv. Bogaerds-Hazenbergh et al., 2022b), en op nogal wat scholen in het voortgezet onderwijs wordt die praktijk in de leesles van het schoolvak Nederlands tot in het examenjaar gecontinueerd (bijv. Canton et al., 2013). Leerlingen hebben daardoor nog weleens een erg smalle visie op tekst-begrip. In de evaluatie van onze lessenserie uit hoofdstuk 6 zagen we daar regel-matig sporen van terug. Zo antwoordde een leerling op de vraag wat tekstbegrip is: “als je vragen over de tekst moet beantwoorden”; en in de slotevaluatie merkte een andere leerling op: “Ik snap niet waarom je het zo moet doen. Je kunt [toch] gewoon de tekst opnieuw lezen als je de vraag niet snapt?” Ook in de lesobservaties kwam het terug. In een klassikale evaluatie na afronding van de lessenserie uit hoofdstuk 6 waren de reacties van vwo 4-leerlingen in eerste instantie gematigd enthousiast (“’t Was best wel anders, mevrouw, maar wat hebben we eraan?”). Een klasgenoot merkte echter op dat hij dankzij de lessenserie voor de eerste keer in een leesles had

nagedacht over hoe hij zelf las, en daarover zou het in de leesles volgens hem eigenlijk altijd moeten gaan. Daarna sloeg de stemming in de klas merkbaar om.

Onze studies in hoofdstuk 5 en 6 onderbouwen dat procesgericht werken in de leesles loont. Echter, zoals we hierboven al aangaven, is het de vraag of docenten niet meer scholing zouden moeten hebben in zo'n benadering in de leesles. Het effectief bespreken van de leerlingmodellen vergt immers een nauwkeurige voorbereiding, een goede kennis van (de effectiviteit van) leesstrategieën, en een gerichte, onderbouwde reflectie op details in zowel beelden als commentaar van de modellen – terwijl leerlingen op voorhand niet overtuigd zijn van de meerwaarde van zo'n procesgerichte benadering. Dat vergt een krachtige didactische hand die lang niet alle docenten Nederlands zomaar tot hun beschikking lijken te hebben.

Bovendien lijken sommige docenten Nederlands zelf evenmin overtuigd van de meerwaarde van een procesgerichte aanpak in leesonderwijs: volgens de docenten van één deelnemende school aan het experiment in hoofdstuk 6 had onze experimentele lessenserie geen duidelijke meerwaarde, omdat ze “geen goede voorbereiding op het Centraal Examen” vormde – wat inderdaad geen doel van de lessenserie was. Net als hun leerlingen, lijken sommige docenten Nederlands een sterk productgerichte benadering van de leesles voor te staan en tekstbegrip weleens te versmallen tot het kunnen beantwoorden van begripsvragen (vgl. een identieke opmerking hierover over leraren in het basisonderwijs in Bogaerds-Hazenberget al., 2022b). Lerarenopleidingen en nascholingsaanbieders doen er daarom goed aan docenten te scholen in een procesgerichte benadering van leesonderwijs en de voordelen die deze benadering heeft.

4.5 *Samenvatting*

In ons onderzoek hebben we op basis van de resultaten in hoofdstuk 2 tot en met 4 de effectiviteit van de tekst met vragen ter discussie gesteld als oefening in de leesles. Er is weinig reden aan te nemen dat leerlingen ermee zelfstandig een tekst leren bestuderen; ze leren ermee waarschijnlijk vooral vragen over een tekst goed te beantwoorden. Ons advies aan de docent Nederlands luidde dan ook: wees terughoudend met de inzet van deze taak in de leesles. Omdat we een andere didactiek wilden ontwikkelen én omdat in een aantal opzichten de huidige didactiek in de leesles bleek aan te sluiten op verouderde inzichten, hebben we in hoofdstuk 5 en 6 een alternatieve lessenserie ontworpen voor de vwo-leesles, waaronder diep begrip, intentionaliteit van tekstbegrip, flexibele, doelgerichte strategie-instructie en observerend-leren-lezen pijlers waren.

In ons onderzoek hebben we daarnaast het belang van betekenisvol, doelgericht lezen onderstreept, met oog voor evaluatie en reflectie – waarbij naar onze overtuiging een breder palet aan teksten en bronnen centraal mag staan. In het algemeen is er in het schoolvak Nederlands een wending te zien naar procesgericht werken (bijv. Bonset & Braaksmā, 2008). In dit slothoofdstuk hebben we de wenselijkheid onderstreept dat deze procesgerichte benadering ook in het leesonderwijs haar weg vindt. Een procesgerichte benadering laat leerlingen nieuwe mogelijkheden

ervaren in hoe ze begrip opbouwen. Om een greep te doen uit de leerlingreacties op onze lessenserie uit hoofdstuk 6: “Ik heb vooral geleerd hoe ik m’n begrip kan verbeteren, bijvoorbeeld door te herhalen in mijn hoofd wat er nou eigenlijk staat”, “het is handig om de grote lijnen in je hoofd te hebben en te herhalen”, “ik moet teksten voor mezelf meer overzichtelijk proberen te maken, door lijstjes in mijn hoofd te maken van bijvoorbeeld verschillende argumenten”, “ik heb geleerd dat je de tekst vaak eerst moet scannen zodat ik daarna beter de belangrijkste dingen inhoud”, “ik moet vaker herlezen, lastige stukken opnieuw lezen” – het is de vraag of het eindeloos oefenen met begripsvragen bij teksten inzichten als deze zou hebben voortgebracht.

5 Een persoonlijke aanbeveling tot slot

In ons onderzoek hebben we ons gericht op het begrip van zakelijke teksten. Zoals we al vaststelden in de inleiding, sloten we hiermee aan bij de standaardpraktijk in het schoolvak Nederlands: in leesparagrafen, in leergangen en in (centrale) toetsing wordt leesvaardigheid teruggebracht tot tekstbegrip bij zakelijke teksten en is zakelijk lezen tamelijk strikt gescheiden van het lezen van niet-zakelijke teksten (CvTE, 2022a, 2022b; Expertgroep Doorlopende Leerlijnen Taal, 2009; etc.).

Persoonlijk denk ik dat het erg goed is deze rigide scheiding tussen zakelijke en niet-zakelijke teksten op te heffen; vanwege het meer persoonlijke karakter van deze aanbeveling verlaat ik nu graag de we-vorm die tot nu toe in dit boek steeds is gebruikt (zie noot 1 in hoofdstuk 1). Deze aanbeveling voeg ik toe als de neerlandicus die ik vanuit allerlei functies de afgelopen twintig jaar ben geweest: letterkundige, redacteur van een onderwijstijdschrift, lesonderzoeker, toetsontwikkelaar – en bovenal docent Nederlands. In strikte zin volgt deze aanbeveling niet noodzakelijkerwijs uit het onderzoek in dit boek; ze dient voor mij persoonlijk echter wel gemaakt.

Vooropgesteld: zoals al in § 1.3 en eerder in dit hoofdstuk is vastgesteld: deze scherpe *a priori*-scheiding stoelt op inzichten uit de neerlandistiek uit de jaren zeventig van de vorige eeuw (vgl. Griffioen, 1975) en ze is om allerlei redenen moeilijk verdedigbaar. In Angelsaksische leergangen worden zakelijke en niet-zakelijke teksten in leesparagrafen vaak afgewisseld en niet, zoals in Nederlandse leergangen vrijwel steeds wel het geval is, in aparte paragrafen ondergebracht (bijv. Fisher & Frey, 2015). De harde scheiding tussen deze twee tekstsoorten is vanuit huidige gangbare leestheorieën zoals de Constructie-Integratietheorie of de Landscaps-theorie ook nauwelijks verdedigbaar: verhalende teksten maken naar huidig inzicht in veel opzichten gebruik van dezelfde principes in betekenisopbouw als zakelijke teksten (vgl. Kintsch, 1998).

Het opheffen van deze scheiding kent naar mijn idee belangrijke voordelen, voor zowel het lezen van niet-zakelijke teksten als het lezen van zakelijke teksten – maar vanuit de kaders van dit boek ga ik hier enkel op voordelen bij zakelijke teksten in. In de zakelijke leesles staan in het schoolvak Nederlands momenteel doorgaans

vooral de meer cognitieve dimensies van tekstbegrip centraal; de meer emotionele, affectieve dimensies van begrip krijgen daarbij in deze lessen weinig aandacht en komen vooral in de literatuurlessen aan bod. Vooral in de literatuurwetenschap heeft men al decennia uitgebreid aandacht voor de potentiële gelaagdheid in betekenisvorming bij teksten én voor de persoonlijke reactie op en interpretatie van teksten (vgl. Brillenburg Wurth & Rigney, 2019). Het opheffen van de scheiding tussen zakelijke en niet-zakelijke teksten kan ruimte bieden om vaker vanuit zakelijke teksten hierop te reflecteren (vgl. ook Van Dijk & Stronks, 2022).

De vraag is immers in hoeverre het effect van zakelijke teksten op lezers met de huidige, louter op cognitie gerichte aanpak momenteel steeds recht gedaan wordt. Stijl en retorica kunnen sterk het begrip van lezers bij een zakelijke tekst beïnvloeden (bijv. Braet, 2007; Schuurs & Breij, 2013); bovendien kunnen stilistische en retorische kenmerken het begrip van teksten hinderen, terwijl ze de vloeiendheid van lezen en het esthetisch plezier tijdens lezen juist bevorderen (bijv. Menninghaus et al., 2015; Reber et al., 2004). Woordkeuze, retorische ingrepen, argumentatie en *framing* sturen de houding van de lezers ten opzichte van een tekstonderwerp op micro- en macroniveau. Om op het aloude Aristotelische inzicht terug te vallen: talige overtuigingskracht bestaat niet alleen uit logos, maar ook uit ethos en pathos (vgl. Braet, 2007; Schuurs & Breij, 2013). Een leesles waarin zakelijke en niet-zakelijke teksten over één onderwerp gebroederlijk naast elkaar figureren, kan ervoor zorgen dat voor deze meer emotionele en affectieve dimensies van begrip óók bij zakelijke teksten in de leesles plaats is.

Zo'n vermenging van zakelijke en niet-zakelijke teksten in de les kan nog een tweede, misschien wel belangrijker voordeel opleveren. Recent werd een pleidooi gehouden om een meer pedagogische visie op het schoolvak Nederlands te ontwikkelen, waarin aan stijlonderwijs een belangrijke rol wordt toegekend in de persoonsvorming van leerlingen (Steenbakkers & De Gloppe, 2020; Steenbakkers et al., 2020):

[E]en kind en adolescent kan niet door een ander kind, of door zichzelf, worden opgevoed tot volwassenheid. Persoonsvorming vraagt om dialogische situaties en om een pedagogiek die weerstand biedt. De ontmoetingen van leerlingen met docenten die zowel vakliefde als volwassenheid durven te belichamen, vormen daarom de kern van pedagogisch vakonderwijs. (Steenbakkers & De Gloppe, 2020, p. 9)

Op dit moment komt persoonsvorming vooral in de lessen literatuur aan de orde, zo stellen Steenbakkers en De Gloppe (2020): via gesprekken over romans en gedichten komen persoonlijke perspectieven op teksten aan bod. Steenbakkers en De Gloppe pleiten ervoor in de taalles in het algemeen meer aandacht te schenken aan stijlperspectieven, aan het perspectief van de ander zoals dat in taal is vervat. Door samen na te denken over het belang van een persoonlijke, effectieve, op de ander gerichte schrijfstijl, kunnen leerlingen via taal reflecteren op zichzelf en op anderen. Steenbakkers en De Gloppe pleiten daarom voor meer op reflectie gericht schrijfstijlonderwijs.

Hun betoog kan echter evenzeer voor het lezen van zakelijke teksten gelden. Op dit moment richt het Nederlandse leesonderwijs zich, zoals gezegd, bij zakelijke

teksten vooral op het feitelijke begrip van teksten (vgl. Gubbels et al., 2019) en te weinig op de beoordeling van teksten door lezers, op evalueren en reflecteren (vgl. Van den Broek et al., 2021) – het lijkt een belangrijke oorzaak waarom Nederlandse leerlingen op de PISA-toets een dalende leesvaardigheid vertonen (Van den Broek et al., 2021; Gubbels, 2020; Gubbels et al., 2019). Leerlingen van nu moeten informatie niet alleen kunnen begrijpen, ze moeten informatie in een breed scala aan bronnen over een onderwerp op waarde kunnen beoordelen, informatie uit meerdere bronnen en documentsoorten kunnen ordenen én tegenstrijdige uitspraken daarin kunnen detecteren, analyseren en verklaren (zie Gubbels et al., 2020; Rooijackers & Gubbels, In voorbereiding; Rooijackers et al., 2021a). *Fake news*, verkapte commerciële belangenbehartiging, (*un*)*gesundes Volksempfinden* liggen in de chaotische, rijkgeschakeerde en multiperspectivistische informatiewereld van nu overal op de loer.

Deze aandacht voor evaluatie en reflectie kán in de les louter (meta)cognitief worden ingevuld. Echter, binnen een breder kader kunnen evaluatie en reflectie in de leesles evengoed instrumenten vormen voor persoonsvorming, zoals voorgestaan door Steenbakkers en De Glopper (2020): docenten kunnen met leerlingen in de leesles nadenken over wat een zakelijke of literaire tekst op inhoudelijk én stilistisch vlak betrouwbaar, geloofwaardig en kwaliteitsrijk maakt. Waarin kun je in een zakelijke tekst herkennen dat deze met kennis van zaken is geschreven? (En waaraan kun je in een fictionele tekst herkennen dat personages geen clichés zijn maar mensen met diepgang?) Binnen zo'n klassengesprek gelden dan geen vaststaande wetten: de persoonlijke reactie van leerlingen op en hun interpretatie van de tekstinhoud en van het stilistisch taalgebruik kan een dankbaar onderwerp van klassikaal gesprek vormen – zowel bij niet-zakelijke (literaire) teksten als bij zakelijke teksten.

In zulke lessen leesvaardigheid worden leerlingen niet alleen in staat gesteld met de moderne rijkgeschakeerde informatiestroom om te gaan, ze betrekken deze lessen vooral ook op zichzelf: hoe gebruik je taal op zo'n manier dat ze op een ander geloofwaardig, betrouwbaar en kwaliteitsrijk overkomt? En andersom: hoe herken je in een tekst of de schrijver ervan vooral zinnige, doordachte zaken te melden heeft of juist vooral allerlei vooroordelen onbesuisd en ongefilterd de wereld heeft in geslingerd? En hoe interpreteer je dan deze woorden van een ander op de juiste manier? Daarover in de leesles, individueel, in groepjes en klassikaal nadenken, via zakelijke én niet-zakelijke teksten, met oog voor persoonlijke verschillen én met respect voor de inhoud van een tekst – daarin ligt, denk ik, de uitdaging van een goede leesles in het schoolvak Nederlands.

LITERATUUR

- Aalsvoort, M. van der. (2016). *Vensters op vakontwikkeling. De betwiste invoering van taalkunde in het examenprogramma Nederlands havo/vwo (1988-2008)*. Dissertatie Radboud Universiteit. Nijmegen: Radboud Universiteit Nijmegen
- Aarnoutse, C., & Verhoeven, L. (2003). Interventieonderzoek op het gebied van begrijpend lezen. *Pedagogische Studiën*, 80(2), 81-91
- Acket, J.M. (1912). Leesonderwijs. *De Nieuwe Taalgids*, 6(2), 77-80.
- Afflerbach, P., Hurt, M., & Cho, B.Y. (2020). Reading comprehension strategy instruction. In D.L. Dinsmore, L.K. Fryer & M.M. Parkinson (red.), *Handbook of Strategies and Strategic Processing* (p. 98-118). New York/Londen: Routledge.
- Afflerbach, P., Pearson, P.D., & Paris, S.G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, 61(5), 364-373.
- Agarwal, P.K., & Roediger III, H.L. (2011). Expectancy of an open-book test decreases performance on a delayed closed-book test. *Memory*, 19(8), 836-852.
- Alderson, J.C. (2005). *Assessing Reading*. Cambridge/New York, etc.: Cambridge University.
- Allen, L.K., & McNamara, D.S. (2020). Defining deep reading comprehension for diverse readers. In E. Birr Moje, P. Afflerbach, P. Enciso & N.K. Lesaux (red.), *Handbook of Reading Research, Volume V* (p. 261-276). New York/Londen: Routledge.
- Almasi, J.F., & Fullerton, S.K. (2012). *Teaching strategic processes in reading*. New York/Londen: Guilford Press.
- Andre, T. (1979). Does answering higher level questions while reading facilitate productive reading? *Review of Educational Research*, 49(2), 280-318.
- Anmarkrud, Ø., Bråten, I., & Strømsø, H.I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences*, 30, 64-76.
- Ardoin, S.P., Binder, K.S., Zawoyski, A.M., Nimocks, E., & Foster, T.E. (2019). Measuring the Behavior of Reading Comprehension Test Takers: What Do They Do, and Should They Do It? *Reading Research Quarterly*, 54(4), 507-529.
- Artelt, C., Schiefele, U., & Schneider, W. (2001). Predictors of reading literacy. *European Journal of Psychology of Education*, 16(3), 363-383.
- Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, 4(3), 359-373.
- Bandura, A., & Walters, R.H. (1977). *Social learning theory. Vol. 1*. Prentice Hall: Englewood cliffs.
- Baron, N.S. (2015). *Words onscreen: The fate of reading in a digital world*. New York: Oxford University.
- Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, 30(4), 441-465.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347-364.

- Bogaerds-Hazenberg, S.T.M., & Evers-Vermeul, J. (2020). Modeling onder de loep, op zoek naar effectieve ingrediënten. In A. Kraal (red.), *Voor u gelezen in 2020: Samenvattingen van actueel internationaal wetenschappelijk onderzoek* (p. 15-23). Rotterdam: Kenniscentrum Begrijpend Lezen.
- Bogaerds-Hazenberg, S., Bergh, H. van den, & Evers-Vermeul, J. (2022a). Onderwijs in tekststructuur. In T. Houtveen & R. van Steensel, *De zeven pijlers van onderwijs in begrijpend lezen* (p. 59-84). Utrecht: Eburon.
- Bogaerds-Hazenberg, S., Evers-Vermeul, J., & Bergh, H. van den. (2022b). What textbooks offer and what teachers teach: an analysis of the Dutch reading comprehension curriculum. *Reading and Writing*, 1497-1523.
- Bogaert, N., Devlieghere, J., Hacquebord, H., et al. (2008). *Aan het werk: adviezen ter verbetering van functionele leesvaardigheid in het onderwijs*. Den Haag: Nederlandse Taalunie.
- Bolt, J., Jensma, M., Jonge, F. de, Klaaijzen, H., Oord, A. van, Pijnappel, B., & Verkade, E. (2018). *KERN Nederlands: onderbouw gymnasium/vwo+, academische leerroute*. Groningen: Boom.
- Bonset, H. (2010). Nederlands in het voortgezet en hoger onderwijs: Hoe sluit dat aan? Deel 1. *Levende Talen Magazine*, 97(3), 16-20.
- Bonset, H. (2020). PISA: wat kunnen onze 15-jarigen eigenlijk niet? Neerlandistiek. Geradpleegd op 28 april 2022 via neerlandistiek.nl/2020/07/pisa-wat-kunnen-onze-15-jarigen-eigenlijk-niet/
- Bonset, H., & Braaksmā, M. (2008). *Het schoolvak Nederlands opnieuw onderzocht. Een inventarisatie van onderzoek van 1997 tot en met 2007*. Enschede: SLO.
- Bonset, H., Boer, M. de, & Ekens, T. (2019). *Nederlands in de onderbouw: een praktische didactiek*. Bussum: Coutinho.
- Bosch, J.H. van den. (1909). Over het oude leesonderwijs. *De Nieuwe Taalgids*, 3, 171-176.
- Bouhnik, D., & Marcus, T. (2006). Interaction in distance-learning courses. *Journal of the American Society for Information Science and Technology*, 57(3), 299-305.
- Bouwer, R., & Koster, M. (2016). *Bringing writing research into the classroom: The effectiveness of Tekster, a newly developed writing program for elementary students*. Dissertatie UU. Utrecht: Universiteit Utrecht.
- Bouwer, R., Goossens, M., Mortier, A.V., Lesterhuis, M., & De Maeyer, S. (2018a). Een comparatieve aanpak voor peer assessment: Leren door te vergelijken. In D. Sluijsmans & M. Segers (red.), *Toetsrevolutie: Naar een feedbackcultuur in het hoger onderwijs* (p. 92-106). Culemborg: Phronese.
- Bouwer, R., Lesterhuis, M., Bonne, P., & De Maeyer, S. (2018b). Applying criteria to examples or learning by comparison: Effects on students' evaluative judgment and performance in writing. *Frontiers in Education*, 3, 1-12.
- Braaksmā, M. (2002). *Observational learning in argumentative writing*. Dissertatie UvA. UvA.
- Braaksmā, M., Rijlaarsdam, G., Bergh, H. van den, & Hout-Wolters, B. van (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction*, 22(1), 1-36.
- Braaksmā, M., Rijlaarsdam, G., Bergh, H. van den, & Hout-Wolters, B. van. (2006). What observational learning in writing courses entails: A multiple case study. *L1 Educational Studies in Language and Literature*, 6(1), 31-62.
- Braasch, J.L., Bråten, I., & McCrudden, M.T. (red.). (2018). *Handbook of multiple source use*. New York/London: Routledge.

- Braet, A. (1979). *Taaldaden: een leergang schriftelijke taalbeheersing voor de bovenbouw havo/vwo*. Twee delen. Groningen: Wolters-Noordhoff.
- Braet, A. (2007). *Retorische kritiek: overtuigingskracht van Cicero tot Balkenende*. Den Haag: Sdu Uitgevers.
- Branden, K. van den. (2003). Leesonderwijs in Vlaanderen: van hoera! naar aha! *Vonk*, 32(3), 12-29.
- Branden, K. van den, Keer, H. van, & Ghesquière, P. (2019). *Sleutels voor effectief begrijpend lezen; Inspiratie voor een eigentijds didactiek in het basisonderwijs*. Brussel: Vlaamse Onderwijsraad.
- Bråten, I., Stadler, M., & Salmerón, L. (2017). The role of sourcing in discourse comprehension. In M.F. Schober, D.N. Rapp, & M.A. Britt (red.), *The Routledge Handbook of Discourse Processes* (p. 141-166). New York/Londen: Routledge.
- Breukink, C. (In voorbereiding.) *Focus op dieper lezen. De effectiviteit van Eye Movement Modeling Examples op proza- en poëzieleesprocessen van havo 4- en vwo 4-leerlingen*. Universiteit Utrecht.
- Briesch, A., Chafouleas, S., & Riley-Tillman, T. (2010). Generalizability and dependability of behavior assessment methods to estimate academic engagement: A comparison of systematic direct observation and direct behavior rating. *School Psychology Review*, 39(3), 408-421.
- Brillenburg Wurth, K., & Rigney, A. (2019). *The Life of Texts; An Introduction to Literary Studies* Amsterdam: AUP.
- Brinke, S. ten. (1976). *The complete mother-tongue curriculum: a tentative survey of all the relevant ways of teaching the mother tongue in secondary education*. Groningen: Wolters-Noordhoff/Longman.
- Britt, M.A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In J. R. Kirby & M. J. Lawson (red.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (p. 276-314). Cambridge: Cambridge University.
- Britt, M.A., Rouet, J.-F., & Durik, A.M. (2018). *Literacy beyond text comprehension: A theory of purposeful reading*. New York/Londen: Routledge.
- Broek, A. van den, Bron, J., Gubbels, J., Gijsel, M., Hoogeveen, M., [et al.]. (2022). *Eindrapportage: Analyse en evaluatie referentieniveaus Nederlandse taal en rekenen*. Nijmegen/Amersfoort: ResearchNed/Expertisecentrum Nederlands/ SLO. Geraadpleegd op 7 juni 2022 via expertisecentrumnederlands.nl/uploads/default/e/i/eindrapportage-analyse-en-evaluatie-referentieniveaus-nederlandse-taal-en-rekenen.pdf
- Broek, P. van den. (2012). Individual and developmental differences in reading comprehension: Assessing cognitive processes and outcomes. In J. Sabatini, E. Albro, & T. O'Reilly (red.), *Measuring up: Advances in how we assess reading ability* (p. 39-58). Lanham etc.: R&L Education.
- Broek, P. van den, & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental re-presentations. *Discourse Processes*, 54(5-6), 360-372.
- Broek, P. van den, Helder, A., Espin, C. & Liende, M. van der. (2021). *Sturen op begrip: Effectief leesonderwijs in Nederland*. Universiteit Leiden.
- Broek, P. van den, Ridsen, K., & Husebye-Hartmann, E. (1995). The role of readers' standard for coherence in the generation of inferences during reading. In R.F. Lorch & E.J. O'Brien (red.), *Sources of coherence in text comprehension* (p. 353-373). Mahwah, NJ: Erlbaum.
- Broek, P. van den, Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of

- reading: Inferences and the on-line construction of a memory representation. In H. van Oostendorp & S.R. Goldman (red.), *The construction of mental representations during reading* (p. 71–98). Erlbaum.
- Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology*, 88(1), 18–37.
- Bruin, E. de. (2018, 26 oktober). Nederlands, wat is dat eigenlijk? [Interview met Marc van Oostendorp.] *NRC Handelsblad*.
- Brunfaut, T., & McCray, G. (2015). *Looking into test-takers' cognitive processes whilst completing reading tasks: a mixed-method eye-tracking and stimulated recall study*. (ARAGs Research Reports). S.l.: British Council.
- Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, 76(4), 683–696.
- Cain, K., Patson, N., & Andrews, L. (2005). Age-and ability-related differences in young readers' use of conjunctions. *Journal of Child Language*, 32(4), 877–892.
- Canton, J., Aler, T., Heemskerk, K., van der Westen, W., & Willemsen, K. (red.). (2013). *Effecten van sturing op discrepantie tussen de cijfers van het centraalexamen: Onderzoek naar de sturing door schoolleiders en de effecten daarvan op het taalonderwijs op havo en vwo en het school-examen bij de talen*. Utrecht: VLLT.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1), 1–9.
- Cataldo, M.G., & Oakhill, J. (2000). Why are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial memory on the ability to locate information in text. *Journal of Educational Psychology*, 92(4), 791–799.
- Catrysse, L., Gijbels, D., Donche, V., De Maeyer, S., Lesterhuis, M., & Van den Bossche, P. (2017). How are learning strategies reflected in the eyes? Combining results from self-reports and eye-tracking. *British Journal of Educational Psychology*, 88(1), 118–137. doi:10.1111/bjep.12181
- Cerdán, R., Pérez, A., Vidal-Abarca, E., & Rouet, J.-F. (2019). To answer questions from text, one has to understand what the question is asking: differential effects of question aids as a function of comprehension skill. *Reading and Writing* 32, 2111–2124.
- Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, 19(1), 13–27.
- Clowes, R.W. (2019). Screen reading and the creation of new cognitive ecologies. *AI & Society*, 34(4), 705–720.
- Cohen, A.D., & Upton, T.A. (2006). Strategies in responding to the new TOEFL reading tasks. *ETS Research Report Series*, 2006(1), i–162.
- Copier, J. (2020, 30 september). Kinderen én leraren moeten weer plezier in lezen krijgen. *Trouw*.
- Cort, F. de. (1868). *Liederen*. Groningen/Antwerpen: J.B. Wolters / J.W. Marchand en Co.
- Couzijn, M.J. (1995). *Observation of writing and reading activities: Effects on learning and transfer*. Dissertatie UvA. Amsterdam: UvA.
- Cozijn, R. (2000). *Integration and inference in understanding causal inferences*. Dissertatie. Tilburg: Faculteit der Letteren, KUB.
- Cozijn, R. (2006). Het gebruik van oogbewegingen in leesonderzoek. *Ephemerides Theologicae Lovanienses*, 28, 220–232.

- Creswell, J.W., & Plano Clark, V.L. (2017). *Designing and conducting mixed methods research*. California: Sage publications.
- Curriculum.nu. (2019). *Toelichting Nederlands; toelichting op het voorstel voor de basis van de herziening van de kerndoelen en eindtermen van de leraren en schoolleiders uit het ontwikkelteam Nederlands*. Geraadpleegd op 24 februari 2022 via curriculum.nu/voorstellen/nederlands/
- Cutting, L.E., & Scarborough, H.S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific studies of reading*, 10(3), 277-299.
- CVEN. (1991). *Eindverslag van de Commissie Vernieuwing Examenprogramma's Nederlandse taal en letterkunde v.w.o. en h.a.v.o.* Den Haag: Sdu Uitgeverij.
- CvTE [College voor Toetsen en Examens]. (2022a). *Nederlands (3F) havo: syllabus Centraal Examen*. Utrecht: CvTE.
- CvTE [College voor Toetsen en Examens]. (2022b). *Nederlands (4F) vwo: syllabus Centraal Examen*. Utrecht: CvTE.
- Damhuis, R., & Blauw, A. de. (2008). High quality interaction in the classroom: a focus for professional learning. *L1-Educational Studies in Language and Literature*, 8(4), 107-124.
- De Milliano, I. (2013). *Literacy development of low-achieving adolescents. The role of engagement in academic reading and writing*. Proefschrift UvA. Amsterdam: UvA.
- De Milliano, I., & Gelderen, A. van. (2020). Taakgericht lezen en schrijven van vmbo'ers; Verschillen in de aanpak van leerlingen en kwaliteit van hun taakuitvoering. *Levende Talen Tijdschrift*, 21(4), 3-15.
- Dera, J., & Doeselaar, N. van. (2022). Leesvoorkeuren van havo- en vwo-leerlingen aan het begin van de bovenbouw; een enquêteonderzoek. *Levende Talen Tijdschrift*, 23(2), 3-13.
- Dera, J., & Steensel, R. van (red.). (2022). Themanummer leesbevordering, laaggeletterdheid en leesonderwijs. *Nederlandse Letterkunde*, 27(1).
- DESI. (2020). *Digital Economy and Society Index (DESI) 2021*. European Commission.
- Diels, P.A. (1938). De triomf van het stil-lezen. *Paedagogische Studiën*, 19, 105-110.
- Dijk, Y. van, & Klaver, M.J. (2022, 18 mei). De leescrisis in het onderwijs; hoe minder je leest, hoe minder je weet. *de Groene Amsterdammer*, 144(20), 4-10.
- Dijk, Y. van, & Stronks, E. (2022). Onderzoekend lezen: geïntegreerd leesonderwijs en de rol van de leerling als betekenisgever. *Nederlandse Letterkunde*, 27(1), 63-83.
- Dis, L.M. van [et al.]. (1962). *Didactische handleiding voor de leraar in de moedertaal*. Amsterdam/Purmerend/Groningen: J.H. Meulenhoff [e.a.].
- Donker, A.S., Boer, H. de, Kostons, D., Ewijk, C.D. van, & Werf, M.P. van der. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, 11, 1-26.
- Dönszelmann, S., Beuningen, C. van, Kaal, A., & Graaff, R. de. (2020). *Handboek Vreemdetalendidactiek: vertrekpunten, vaardigheden, vakinhoud*. Bussum: Coutinho.
- Drop, W. (1979). *Instrumentele tekstanalyse ten dienste van samenvatten, opstellen van begripsvragen en tekstverbeteren*. Groningen: Wolters-Noordhoff.
- Drop, W., & Vries, J.H.L. de. (1974). *Taalbeheersing: handboek voor taalhantering*. Groningen: Wolters-Noordhoff.
- Duchowski, A.T. (2007). *Eye tracking methodology: Theory and practice*. Londen: Springer.
- Duke, N.K., & Pearson, P.D. (2002). *What research has to say about reading instruction*. Newark, DE: International Reading Association.
- Duke, N.K., Pearson, P.D., Strachan, S.L., & Billman, A.K. (2011). Essential elements of fos-

- tering and teaching reading comprehension. In: J. Samuels, & A.E. Farstrup (red.), *What research has to say about reading instruction* (p. 51-93). Newark: International Reading Association.
- Dunlosky, J., & Rawson, K.A. (2005). Why does rereading improve metacomprehension accuracy? Evaluating the levels-of-disruption hypothesis for the reading effect. *Discourse Processes, 40*, 37-55.
- Durkin, D. (1978). What classroom observation reveals about reading comprehension instruction. *Reading Research Quarterly 14*, 481-533.
- Edmonds, W.A., & Kennedy, T.D. (2016). *An applied guide to research designs: Quantitative, qualitative, and mixed methods*. Londen: Sage.
- Elving-Heida, K. (2019). *Effectieve leeractiviteiten voor het schrijfonderwijs in havo 4*. (Proefschrift UU). Utrecht: LOT.
- Embretson, S.E., & Wetzel, C.D. (1987). Component latent trait models for paragraph comprehension tests. *Applied Psychological Measurement, 11*(2), 175-193.
- Ent, W. van den [et al.]. (1941). *Het onderwijs in de Nederlandse taal en letterkunde op de middelbare school: rapport van de commissie ter bestudering van de middelen tot verbetering van het onderwijs in het Nederlands [etc.]*. 's-Gravenhage: z.u.
- Evers-Vermeul, J., Heesters, K., Schuurs, U., & Rooijackers, P. (2022). *Opbrengstenvanhetproject 'UitbreidingvanvraagtypenvoorhetleesvaardigheidsexamenNederlandshavoenvwo'*. S.l.: CvTE/Cito. Geraadpleegd op 7 juni 2022 via examenblad.nl/nieuws/20220519/project-leesvaardigheid-nederlands/2022/vwo?regime=hflinks&horizon=vhi1nc4phqya
- ExamenOverzicht. (2020). *Oefenboek Nederlands Examenstof 2020-2021*. Z.p.: Examen-Overzicht.
- Expertgroep Doorlopende Leerlijnen Taal. (2008). *Over de drempels met taal*. Enschede: SLO.
- Expertgroep Doorlopende Leerlijnen Taal. (2009). *Een nadere beschouwing: over de drempels met taal en rekenen*. Enschede: SLO.
- Fang, Z. (2016). Teaching close reading with complex texts across content areas. *Research in the Teaching of English, 51*(1), 106-116.
- Ferrer, A., Vidal-Abarca, E., Serrano, M.Á., & Gilabert, R. (2017). Impact of text availability and question format on reading comprehension processes. *Contemporary Educational Psychology, 51*, 404-415.
- Fisher, D., & Frey, N. (2013). *Gradual release of responsibility instructional framework*. IRA E-ssentials, 1-8.
- Fisher, D., & Frey, N. (2015). *Text-dependent questions, grades K-5: Pathways to close and critical reading*. Thousand Oaks: Corwin.
- Fisher, D., Frey, N., & Lapp, D. (2008). Shared readings: Modeling comprehension, vocabulary, text structures, and text features for older readers. *The Reading Teacher, 61*(7), 548-556.
- Fox, E., & Alexander, P.A. (2017). Text and comprehension: A retrospective, perspective, and prospective. In S. E. Israel (red.), *Handbook of research on reading comprehension* (p. 335-352). New York/Londen: Guilford Press.
- Frank, H., Grezel, J.E., Kooiman, M., Steenbergen, W., Vos, M., & Zetstra, H. (2019). *Nieuw Nederlands: 4-5-6 vwo*. Groningen/Utrecht: Noordhoff.
- Gelderen, A. van, & Schooten, E. van. (2011). *Taalonderwijs; een kwestie van ontkavelen: Openbare les in duplo*. Rotterdam: Rotterdam University Press.
- Gellert, A.S., & Elbro, C. (2013). Cloze tests may be quick, but are they dirty? Development

- and preliminary validation of a cloze test of reading comprehension. *Journal of Psycho-educational Assessment*, 31(1), 16-28.
- Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during Web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, 21(2), 220-231.
- Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H.I. (2010). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology*, 31(1), 30-68.
- Gil, L., Martinez, T., & Vidal-Abarca, E. (2015). Online assessment of strategic reading literacy skills. *Computers & Education*, 82, 50-59.
- Glenberg, A.M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, 15(1), 84-93.
- Goldman, S.R. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes*, 23(3), 357-398.
- Goldstein, H. (1987). *Multilevel models in education and social research*. Oxford: Oxford University.
- Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371-395.
- Graham, S., & Harris, K.R. (1994). Cognitive strategy instruction: Issues and guidelines in conducting research. In S. Vaughn & C. Bos (red.), *Research issues in learning disabilities: Theory, methodology, assessment, and ethics* (p. 146-160). New York: Springer.
- Graham, S., & Hebert, M. (2011). Writing-to-read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, 81, 710-744.
- Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescent middle and high school*. New York, NY: Carnegie Corporation of New York.
- Grenner, E., Åkerlund, V., Asker-Årmason, L., Weijer, J. van de, Johansson, V., & Sahlén, B. (2020). Improving narrative writing skills through observational learning: a study of Swedish 5th-grade students. *Educational Review*, 72(6), 691-710. doi.org/10.1080/00131911.2018.1536035
- Grezel, J.E. (2014). Vaardigheid of kunstje? Eindexamen Nederlands onder vuur. *Onze Taal*, 83(5), 109-115.
- Griffioen, J. (1975). *Zeggen-schap: grondslagen en een uitwerking van een didactiek van het Nederlands in het voortgezet onderwijs*. Groningen: Wolters-Noordhoff.
- Groenendijk, T., Janssen, T., Rijlaarsdam, G., & Bergh, H. van den. (2011). The effect of observational learning on students' performance, processes, and motivation in two creative domains. *British Journal of Educational Psychology*, 83(1), 3-28.
- Groenendijk, T., Janssen, T., Rijlaarsdam, G., & Bergh, H. van den. (2013). Learning to be creative. The effects of observational learning on students' design products and processes. *Learning and Instruction*, 28, 35-47. <https://doi.org/10.1016/j.learninstruc.2013.05.001>
- Gubbels, J. (2020). Moeite met evalueren en reflecteren: wat betekent dat? *Didactiek Nederlands - Handboek*. Geraadpleegd op 5 juli 2021 via didactieknederlands.nl/handboek/2020/12/moeite-met-evalueren-en-reflecteren-wat-betekent-dat
- Gubbels, J., Langen, A. van, Maassen, N., & Meelissen, M. (2019). *Resultaten PISA-2018 in vogelvlucht*. Enschede: Universiteit Twente.
- Gubbels, J., Netten, A., & Verhoeven, L. (2017). *Vijftien jaar leesprestaties in Nederland: PIRLS-2016*. Nijmegen: Expertisecentrum Nederlands/ Radboud Universiteit.
- Guthrie, J.T., Van Meter, P., McCann, A.D., Wigfield, A., Bennett, L., Poundstone, C.C.,

- Rice, M.E., Faibisch, F.M., Hunt, B., & Mitchell, A.M. (1996). Growth of literacy engagement: Changes in motivations and strategies during concept-oriented reading instruction. *Reading Research Quarterly*, 31(3), 306-332.
- Haas, E. (1995). *Op de juiste plaats: de opkomst van de bedrijfs- en schoolpsychologische beroepspraktijk in Nederland*. Hilversum: Verloren.
- Halamish, V., & Elbaz, E. (2020). Children's reading comprehension and metacomprehension on screen versus on paper. *Computers & Education*, 145, 103737.
- Hamilton, R.J. (1985). A framework for the evaluation of the effectiveness of adjunct questions and objectives. *Review of Educational Research*, 55, 47-85.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York/Londen: Routledge.
- Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, 66(2), 99-136.
- Hayles, N.K. (2012). *How we think: Digital media and contemporary technogenesis*. Chicago: University of Chicago Press.
- Heemskerk, K. (2013, 23 mei). Bij Nederlands leer je nu vooral veel trucjes. *NRC Handelsblad*.
- Hendrix, T., & Hulshof, H. (1994). *Leesvaardigheid Nederlands: Omgaan met zakelijke teksten*. Bussum: Coutinho.
- Hermida, D. (2009). The importance of teaching academic reading skills in first-year university courses. *International Journal of Research and Review*, 3, 20-30.
- Higgs, K., Magliano, J.P., Vidal-Abarca, E., Martínez, T., & McNamara, D.S. (2017). Bridging skill and task-oriented reading. *Discourse Processes*, 54(1), 19-39.
- Hill, L.G., & Betz, D.L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, 26(4), 501-517.
- Holland, N. (1975). *5 Readers Reading*. New Haven, CT: Yale University Press.
- Holmqvist, K., & Andersson, R. (2017). *Eye Tracking: A comprehensive guide to methods, paradigms and measures*. CreateSpace Independent Publishing Platform.
- Hoogeveen, M. (2018). *Het leesonderwijs in primair en voortgezet onderwijs: een stand van zaken en curriculaire aanbevelingen*. Enschede: SLO.
- Hoogeveen, M., & Bonset, H. (1998). *Het schoolvak Nederlands onderzocht*. Leuven/Apeldoorn: Garant.
- Houtveen, A.A.M., Steensel, R. van, Rie, S. de la. (2019). *De vele kanten van leesbegrip: Literatuurstudie naar onderwijs in begrijpend lezen in opdracht van het Nationaal Regieorgaan Onderwijsonderzoek en de Inspectie van het Onderwijs*. Rotterdam: Erasmus Universiteit Rotterdam.
- Houtven, T. van, Peters, E., & El Morabit, Z. (2010). Hoe staat het met de taal van studenten? Exploratieve studie naar begrijpend lezen en samenvatten bij instromende studenten in het Vlaamse hoger onderwijs. *Levende Talen Tijdschrift*, 11(3), 29-44.
- Hulshof, H., Kwakernaak, E., & Wilhelm, F. (2015). *Geschiedenis van het talenonderwijs in Nederland: onderwijs in de moderne talen van 1500 tot heden*. Groningen: Passage.
- Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychologica*, 73(3), 259-280.
- Hyönä, J., Lorch Jr, R.F., & Kaakinen, J.K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, 94(1), 44-55.
- Inspectie van het Onderwijs. (2019). *De staat van het onderwijs 2018. Onderwijsverslag over 2017/2018*. Utrecht: Inspectie van het Onderwijs.

- Inspectie van het Onderwijs. (2022). Factsheet leesvaardigheid einde (s)bo 2020-2021: technische toelichting. Geraadpleegd op 22 juni 2022 via onderwijsinspectie.nl/binaries/onderwijsinspectie/documenten/rapporten/2022/06/08/technische-toelichting-factsheet-peil.leesvaardigheid-einde-s-bo-2020-2021/Technische+toelichting+Factsheet+Leesvaardigheid+juni+2022.pdf
- Jackson, M.D., & McClelland, J.L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology*, 108(2), 151-181.
- Jarodzka, H., Gog, T. van, Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, 25, 62-70.
- Johnston, P. (1984). Prior Knowledge and Reading Comprehension Test Bias. *Reading Research Quarterly*, 19(2), 219-239.
- Jones, R., & Hafner, C. (2012). *Understanding digital literacies: A practical introduction*. New York/Londen: Routledge.
- Kalmthout, T. van. (2002). Vormend lezen: opvattingen over de taak van het literaire leesonderwijs, ca. 1890-1940. *Nederlandse Letterkunde*, 7(2), 101-132.
- Kalshoven, F. (2019, 6 december). Een kwart van de tieners is laaggeletterd. Er is verdorie een leescrisis. *de Volkskrant*.
- Kamalski, J. (2007). *Coherence marking, comprehension and persuasion: On the processing and representation of discourse*. Dissertatie UU. LOT. Utrecht University Repository. <https://dspace.library.uu.nl/handle/1874/22617>
- Keehnen, T., Braaksma, M., & Boer, M. de. (2015). Leren door zien lezen. Observerend leren bij leesvaardigheid in 3 vwo. *Levende Talen Tijdschrift*, 16(1), 34-41.
- Keenan, J.M., Betjemann, R.S., & Olson, R.K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281-300.
- Kellogg, R.T. (1994). *The psychology of writing*. Oxford: Oxford University.
- Kintsch, W. (1998). *Comprehension: a paradigm of cognition*. Cambridge: Cambridge University.
- Kintsch, W., & Dijk, T.A. van. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363-394.
- Kirschner, P.A. (2017). *Het voorbereiden van leerlingen op (nog) niet bestaande banen*. Arnhem: NVSP Innovatief in Werk.
- Kleijn, S. (2018). *Clozing in on readability: How linguistic features affect and predict text comprehension and on-line processing*. Utrecht: LOT. Dissertatie UU.
- Knoop-van Campen, C.A., Kok, E., Doornik, R. van, Vries, P. de, Immink, M., Jarodzka, H., & Gog, T. van. (2021). How teachers interpret displays of students' gaze in reading comprehension assignments. *Frontline Learning Research*, 9(4), 116-140.
- Kraal, A. (2020). *Different readers, different texts, different processes: the effects of reader and text properties on text processing*. Dissertatie. Universiteit Leiden.
- Kramer, W. (1938). De triomf van het stil-lezen en de gevaren van dien. *De Nieuwe Taalgids* 32, 273-280.
- Kreeft, H., & Luijten, T. (1979). Teksten met vragen: het gebruik van tekstonafhankelijke vraagssystemen bij leesvaardigheid. In *DCN-cahier nr. 8: omgaan met teksten in de klas* (Groningen: Wolters-Noordhoff), 59-80.
- Kreft, I.G., & Leeuw, J. de. (2007). *Introducing multilevel modeling*. Londen: Sage.
- Kuijpers, C., Leij, A. van der, Been, P., Leeuwen, T. van, Keurs, M. ter, Schreuder, R., & Bos,

- K.P. van den. (2003). Leesproblemen in het voortgezet onderwijs en de volwassenheid. *Pedagogische Studiën*, 80(4), 272-287.
- Kuperman, V., & Van Dyke, J.A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, 65(1), 42-73. <https://doi.org/10.1016/j.jml.2011.03.002>
- Lam, T.C., & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, 24(1), 65-80.
- Land, J. (2009). *Zwakke lezers, sterke teksten? Effecten van tekst- en lezerskenmerken op het tekstbegrip en de tekstwaardering van vmbo-leerlingen*. Delft: Eburon.
- Langeveld, J. (1934). *Taal en denken: een theoretische en didactische bijdrage tot het onderwijs in de moedertaal op de middelbare school, inzonderheid tot dat der grammatika*. Groningen/Den Haag/Batavia: J.B. Wolters.
- Lapp, D., Moss, B., Grant, B., & Johnson, K. (2018). *Close reading: werken aan dieper tekstbegrip in het basisonderwijs*. Soest: Pica.
- Leest, J. (1932). *Het voortgezet onderwijs in de moedertaal*. Groningen/Den Haag/Batavia: J.B. Wolters.
- Leeuw, L. de. (2015). *Understanding reading comprehension processes across the primary grades*. Dissertatie RUN. Nijmegen: Radboud Universiteit.
- Leeuw, L. de, Segers, E., & Verhoeven, L. (2016). The effect of student-related and text-related characteristics on student's reading behaviour and text comprehension: An eye movement study. *Scientific Studies of Reading*, 20(3), 248-263.
- Leeuw, B. van der, Meestringa, T., Silfhout, G. van, Smit, J., Hoogeveen, M., Prenger, J., Langberg, M., & Jansma, N. (2017). *Nederlands; Vakspecifieke trendanalyse 2017*. Enschede: SLO.
- Liaw, S.S. (2008). Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the Blackboard system. *Computers & Education*, 51(2), 864-873.
- Linderholm, T., & Broek, P. van den. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, 94(4), 778.
- Linthorst, T.R., & Glopper, K. de. (2015). De didactiek van begrijpend lezen in het voortgezet onderwijs: lesobservaties bij Nederlands en zaakvakken. *Pedagogische Studiën*, 92(2), 150-166.
- Liu, Z. (2005). Reading behavior in the digital environment. *Journal of Documentation*, 61(6), 700-712.
- Loo, J. van der. (2022). *Mastering The Art of Academic Writing: Comparing the Effectiveness of Observational Learning and Learning by Doing*. Dissertatie. Universiteit van Tilburg.
- Lorch, R.F., Lorch, E.P., & Klusewitz, M.A. (1993). College students' conditional knowledge about reading. *Journal of Educational Psychology*, 85(2), 239-252.
- Lorch, R.F., Lorch, E.P., & Matthews, P.D. (1985). On-line processing of the topic structure of a text. *Journal of Memory and Language*, 24(3), 350-362.
- Lorch, R.F., Lorch, E.P., & Mogan, A.M. (1987). Task effects and individual differences in on-line processing of the topic structure of a text. *Discourse Processes*, 10(1), 63-80.
- LTN [Levende Talen Nederlands]. (2019a). *Evaluatie Examens Nederlands 2019*. Z.p.: Levende Talen Nederlands. Geraadpleegd op 10 juli 2021 via lerarennederlands.nl/wp-content/uploads/2020/03/Evaluatie-Examens-Nederlands-2019-met-voorblad.pdf

- LTN [Levende Talen Nederlands]. (2019b). *Advies curriculum Nederlands in de bovenbouw van het voortgezet onderwijs*. Z.p.: Levende Talen Nederlands. Geraadpleegd op 10 juli 2021 via levendetalen.nl/wp-content/uploads/2019/09/LT-advies-curriculum-Nederlands-2019.pdf
- Lysynchuk, L.M., Pressley, M., & Vye, N.J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. *The Elementary School Journal*, 90(5), 469-484.
- Magliano, J.P., McCrudden, M.T., Rouet, J.-F., & Sabatini, J. (2018). The modern reader: Should changes to how we read affect research and theory? In M.F. Schober, D.N. Rapp, & M.A. Britt (red.), *The Routledge handbook of discourse processes* (p. 343-361). New York: Routledge.
- Magliano, J.P., Millis, K., Ozuru, Y., & McNamara, D.S. (2007). A multidimensional framework to evaluate reading assessment tools. In D.S. McNamara (red.), *Reading comprehension strategies: Theories, interventions, and technologies* (p. 107-136). Mahwah, NJ: Erlbaum.
- Maki, R.H. (1998). Test predictions over text material. In D.J. Hacker, A.C. Graesser, & J. Dunlosky (red.), *Metacognition in educational theory and practice* (p. 117-144). New York/Londen: Routledge.
- Maki, R.H., & Berry, S.L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(4), 663-679.
- Maki, R.H., Shields, M., Wheeler, A.E., & Zachilli, T.L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, 97(4), 723-731.
- Máñez, I., Vidal-Abarca, E., Kendeou, P., & Martínez, T.. (2019). How do students process complex formative feedback in question-answering tasks? A think-aloud study. *Metacognition Learning*, 14, 65-87. <https://doi.org/10.1007/s11409-019-09192-w>
- Mangen, A., Walgermo, B.R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, 58, 61-68.
- Marlen, T. van. (2019). *Looking Through the Teacher's Eyes: Effects of Eye Movement Modeling Examples on Learning to Solve Procedural Problems*. Dissertatie UU. Universiteit Utrecht.
- Martínez, R.A., Del Bosch, M.M., Herrero, M.H.P., & Nuño, A.S. (2007). Psychopedagogical components and processes in e-learning. Lessons from an unsuccessful on-line course. *Computers in human behavior*, 23(1), 146-161.
- McCrudden, M.T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review*, 19(2), 113-139.
- McCrudden, M.T., Magliano, J.P., & Schraw, G. (2010). Exploring how relevance instructions affect personal reading intentions, reading goals and text processing: A mixed methods study. *Contemporary Educational Psychology*, 35(4), 229-241.
- McCrudden, M.T., Magliano, J.P., & Schraw, G. (2011). Toward an integrated view of relevance in text comprehension. In M.T. McCrudden, J.P. Magliano, & G. Schraw (red.), *Text relevance and learning from text* (p. 395-414). Charlotte, NC: Information Age.
- McNamara, D.S. (red.) (2007). *Reading comprehension strategies: Theory, interventions, and technologies*. Mahwah, NJ: Erlbaum.
- McNamara, D.S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22(3), 247-288.

- McNamara, D.S., & Magliano, J.P. (2009a). Toward a comprehensive model of comprehension. In B.H. Ross (red.), *The Psychology of Learning and Motivation, Vol. 51* (p. 297-384). Amsterdam: Academic Press.
- McNamara, D.S., & Magliano, J.P. (2009b). Self-explanation and metacognition. In D.J. Hacker et al. (red.), *Handbook of metacognition in education* (p. 60-81). New York/Londen: Routledge.
- Meesterschapsteam. (2019). *Perspectieven op taal, taalgebruik en literatuur als basis voor nieuwe eindtermen*. Geraadpleegd op 24 februari 2022 via nederlands.vakdidactiekgw.nl/wp-content/uploads/sites/4/2019/11/Perspectieven_als_basis_voor_eindtermen_NL_LINEAIR_eindversie181219.pdf
- Meestringa, T., & Ravesloot, C. (2012). *Het schoolexamen Nederlands op havo en vwo: Verslag van een digitale enquête*. Enschede: SLO.
- Menninghaus, W., Bohrn, I.C., Knoop, C.A., Kotz, S.A., Schlotz, W., & Jacobs, A.M. (2015). Rhetorical features facilitate prosodic processing while handicapping ease of semantic comprehension. *Cognition*, 143, 48-60.
- Merckx, P., Flier, E., & Alers, R. (2018). *Op niveau tweede fase: leeropdrachtenboek 4 vwo*. Amersfoort: ThiemeMeulenhoff.
- Merrill, M.D. (2002). First principles of instruction. *Educational Technology, Research and Development*, 50(3), 43-59.
- Miller, B.W. (2015). Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist*, 50(1), 31-42.
- Meulen, F. van der. (2015). Welke methode verbetert leesvaardigheid het meest? Een vergelijkend onderzoek naar vier lesmethoden in 4 vwo. *Levende Talen Tijdschrift*, 16(1), 14-23.
- Ministerie van Onderwijs, Cultuur & Wetenschappen. (2020, 13 oktober). Kamerbrief 'Stand van zaken leesoffensief'. Geraadpleegd op 24 februari 2022 via rijksoverheid.nl/documenten/kamerstukken/2020/10/13/kamerbrief-stand-van-zaken-leesoffensief
- Ministerie van Onderwijs, Cultuur & Wetenschappen. (2022, 12 mei). Kamerbrief over Masterplan basisvaardigheden. Geraadpleegd op 7 september 2022 via rijksoverheid.nl/documenten/kamerstukken/2022/05/12/kamerbrief-masterplan-basisvaardigheden
- Moeken, N., Kuiken, F., & Welie, C. (2015). *SALEVO: Samenwerkend lezen in het voortgezet onderwijs*. Amsterdam: Universiteit van Amsterdam. Geraadpleegd via onderwijsdata-bank.nl/94373/samenwerkend-lezen-in-het-voortgezet-onderwijs
- Moormann, J.G.M. (1936). *De moedertaal: een didactiek voor het middelbaar (en lager) onderwijs*. Nijmegen/Utrecht: Dekker & Van de Vegt.
- Mortel, K. van de, & Ballering, C. (2022). *Verdiepend lezen; Leerlingen begeleiden bij het begrijpend lezen van complexe teksten*. CPS.
- Mross, E.F. (1989). *Macroprocessing in expository text comprehension*. Dissertatie. University of Colorado: Boulder.
- Mullis, I.V.S., Martin, M.O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading*. TIMSS & PIRLS International Study Center.
- NAEP [National Assessment of Educational Progress]. (2017). *Evaluation of the achievement levels for mathematics and reading on the National Assessment of Educational Progress*. National Academies Press.
- Narvaez, D., Broek, P. van den, & Barron-Ruiz, A. (1999). The influence of reading purpose on inference-generation and comprehension in reading. *Journal of Educational Psychology*, 91(3), 488-496.

- Nederlands Nu! & Sectiebestuur Nederlands Levende Talen. (2018). *Advies Examens Nederlands [aangeboden aan curriculum.nu]*. S.l.
- NJI [Nederlands Jeugd Instituut]. (2022). Onderzoek naar corona en onderwijs. Geraadpleegd op 22 juni 2022 via nji.nl/coronavirus/onderzoek-naar-corona-en-onderwijs
- Norris, S.P. (1990). Effect of eliciting verbal reports of thinking on critical thinking test performance. *Journal of Educational Measurement*, 27(1), 41-58.
- Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, 42(1), 188-204.
- O'Reilly, T., & McNamara, D.S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional 'high-stakes' measures of high school students' science achievement. *American Educational Research Journal*, 44(1), 161-196.
- O'Reilly, T., Feng, D.G., Sabatini, D.J., Wang, D.Z., & Gorin, D.J. (2018). How do people read the passages during a reading comprehension test? The effect of reading purpose on text processing behavior. *Educational Assessment*, 23(4), 277-295.
- Ockenburg, L. van. (2022). *Synthesis writing: teaching high schools how to read, plan, draft and revise*. Dissertatie. Amsterdam: UvA.
- OECD. (2019). *PISA 2018: Assessment and Analytical Framework*. Paris: PISA, OECD. doi.org/10.1787/b25efab8-en
- Okkinga, M., Steensel, R. van, Gelderen, A.J. van, Schooten, E. van, Slegers, P.J., & Arends, L.R. (2018). Effectiveness of reading-strategy interventions in whole classrooms: A meta-analysis. *Educational Psychology Review*, 30(4), 1215-1239.
- Oller, J. W., & Jonz, J. (1994). *Cloze and Coherence*. Associated University.
- Onderwijsinspectie. (2022). De gevolgen van de coronacrisis zijn voor vo-leerlingen ongelijk verdeeld. Geraadpleegd op 22 juni 2022 via onderwijsinspectie.nl/onderwerpen/corona-onderzoeken/
- Onderwijsraad. (2019). *Doorgesloten differentiatie in het onderwijsstelsel; stand van educatief Nederland 2019*. Den Haag: Onderwijsraad.
- Onderwijsraad & Raad voor Cultuur. (2019). *Lees! Een oproep tot een leesoffensief*. Raad voor Cultuur/Onderwijsraad.
- Onna, M. van, Straat, H., & Alberts, R. (2014). *Verslag ijkingsonderzoek centrale examens Nederlands VO 2014 aan de referentiesets en referentieniveaus taal*. Arnhem: Cito.
- Oostdam, R., & Bimmel, P. (1996). De kloof tussen theorie en praktijk; Experimenteel onderzoek naar training en transfer van leesstrategieën. *Levende Talen*, 83(511), 332-337.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D.S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25(4), 399-438.
- Ozuru, Y., Briner, S., Kurby, C.A., & McNamara, D.S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology*, 67(3), 215-222.
- Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D.S. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods*, 40(4), 1001-1015.
- Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction*, 1(2), 117-175.

- Paris, S.G., Lipson, M.Y., & Wixson, K.K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8(3), 293-316.
- Parreren, C.F. van, Velema, E., & Peeck, J. (1969). *Effectief studeren*. Utrecht/Antwerpen: het Spectrum.
- Payne, S.J., & Reader, W.R. (2006). Constructing structure maps of multiple on-line texts. *International Journal of Human-Computer Studies*, 64(5), 461-474.
- Pearson, P.D., & Cervetti, G. (2013). The psychology and pedagogy of reading processes. In W. Reynolds & G. Miller (red.), *Handbook of psychology: Educational psychology*, 7 (p. 507-554). New York: John Wiley & Sons.
- Pearson, P.D., & Hamm, D.N. (2005). The assessment of reading comprehension: A review of practices-past, present, and future. In S.G. Paris & S.A. Stahl (red.), *Children's reading comprehension and assessment* (p. 13-69). Routledge.
- Pearson, P.D., Palincsar, A.S., Biancarosa, G., & Berman, A.I. (2020). *Reaping the Rewards of the Reading for Understanding Initiative*. National Academy of Education.
- Pereira, C., & Nicolaas, M. (2019). *Effectief onderwijs in begrijpend lezen*. S.l.: Nederlandse Taalunie.
- Perfetti, C.A. (1985). *Reading ability*. Oxford: Oxford University.
- Perfetti, C.A. (1997). Sentences, individual differences, and multiple texts: Three issues in text comprehension. *Discourse Processes*, 23(3), 337-355.
- Perfetti, C.A., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22-37.
- Perfetti, C.A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M.J. Snowling & C. Hulme (red.), *The science of reading: A handbook* (p. 227-247). Malden: Blackwell.
- Perfetti, C.A., Rouet, J.-F., & Britt, M.A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S.R. Goldman (red.), *The construction of mental representations during reading* (p. 88-109). Mahwah, NJ: Erlbaum.
- Piper, A. (2012). *Book was there: Reading in electronic times*. Chicago: University of Chicago.
- Popham, W.J. (2001). *The truth about testing: An educator's call to action*. Beauregard St. Alexandria: ASCD.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale: Erlbaum.
- Pressley, M., Brown, R., El-Dinary, P.B., & Afflerbach, P. (1995). The comprehension instruction that students need: instruction fostering constructively responsive reading. *Learning Disabilities Research & Practice*, 10(4), 215-224.
- Pronk-van Eunen, M., & Vos, B. de. (2018). Lezen met de leessandwich: Een kansrijke didactiek bij Nederlands en mvt. *Levende Talen Magazine*, 105(2), 10-14.
- Quené, H., & Bergh, H. van den. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43(1-2), 103-121.
- Raad voor Nederlandse Taal en Letteren. (2015). *Vaart met taalvaardigheid: Nederlands in het hoger onderwijs*. Den Haag: Taalunie.
- Raedts, M., Rijlaarsdam, G., Van Waes, L., & Daems, F. (2006). Observational learning through video-based models: impact on students' accuracy of self-efficacy beliefs, task knowledge and writing performances. In G. Rijlaarsdam, P. Boscolo & S. Hidi (red.), *Writing and motivation* (p. 219-238). Oxford: Elsevier.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.

- Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457-1506.
- Reber, R., Schwarz, N., & Winkelman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4), 364-382.
- Riemens-Feteris, E.S.F. (1971). De problematiek van de tekstbehandeling als examenonderdeel. *Levende Talen*, 333-339.
- Rijksoverheid. (2022). Curriculum voor de toekomst. Geraadpleegd op 22 juni via [rijks-overheid.nl/onderwerpen/toekomst-onderwijs/toekomstgericht-curriculum](https://overheid.nl/onderwerpen/toekomst-onderwijs/toekomstgericht-curriculum)
- Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Kieft, M., Broekkamp, H., & Bergh, H. van den. (2005). Psychology and the teaching of writing in 8000 and some words. In *BJEP Monograph Series II, Number 3-Pedagogy-Teaching for Learning* (Vol. 127, No. 153, p. 127-153). British Psychological Society.
- Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Raedts, M., Steendam, E. van, Toorenaar, A., & Bergh, H. van den. (2008). Observation of peers in learning to write. Practise and research. *Journal of Writing Research*, 1(1), 53-83.
- Robben, T. (2008). Over de drempels met taal en rekenen[: Interview met G. Rijlaarsdam & H. van den Bergh]. *Levende Talen Magazine*, 95(5), 5-8.
- Robben, T. (2012). Vrolijk lezen met meer hersens: de mindmap als didactisch instrument voor de ontwikkeling van leesvaardigheid. *Levende Talen Magazine*, 99(4), 4-8.
- Robinson, F.P. (1946). *Effective Study*. New York: Harper.
- Rogiers, A. (2019). *Illuminating text-learning in secondary education: measuring and fostering strategy use and knowledge acquisition*. Dissertatie. Universiteit Gent.
- Rooijackers, P., & Gubbels, J. (In voorbereiding.) Leesbegrip. In J. Dera, J. Gubbels, J. van der Loo, & J. van Rijjt (red.), *Vaardig met vakinhoud: handboek vakdidactiek Nederlands*. Bussum: Coutinho.
- Rooijackers, P., Silfhout, G. van, & Bergh, H. van den. (2021a). Met zijwieltjes leren fietsen in de Dorpsstraat: Waarom de leestaken in de les Nederlands vaak inhoudelijk tekortschieten. *Levende Talen Magazine*, 108(2), 4-9.
- Rooijackers, P., Silfhout, G. van, & Bergh, H. van den. (2021b). Timmeren met oude spijkers en zonder hout: Waarom leergangen Nederlands nu niet echt aan strategie-instructie doen. *Levende Talen Magazine*, 108(1), 18-23.
- Ros, B., Gelderen, A. van, Glopper, K. de, & Steensel, R. van. (2021). *Leer ze lezen: praktische inzichten uit onderzoek voor leraren basisonderwijs*. Meppel: Ten Brink.
- Rosenblatt, L.M. (1938). *Literature as exploration*. New York/Londen: D. Appleton-Century Co.
- Rouet, J.-F. (2006). *The skills of document use: From text comprehension to Web-based learning*. New York/Londen: Routledge.
- Rouet, J.-F., & Britt, M.A. (2011). Relevance processes in multiple document comprehension. In M. McCrudden, J. Magliano, & G. Schraw (red.), *Text relevance and learning from text* (p. 19-52). Information Age Publishing.
- Rouet, J.-F., Britt, M.A., & Durik, A.M. (2017). RESOLV: Readers' Representation of Reading Contexts and Tasks. *Educational Psychologist*, 52(3), 200-215. DOI: 10.1080/00461520.2017.1329015
- Rouet, J.-F., Vidal-Abarca, E., Erbou, A.B., & Millogo, V. (2001). Effects of information search tasks on the comprehension of instructional text. *Discourse Processes*, 31(2), 163-186.

- Rupp, A.A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23(4), 441-474.
- Sabatini, J., O'Reilly, T., Wang, Z., & Dreier, K. (2018). Scenario-Based Assessment of Multiple Source Use. In J.L. Braasch, I. Bråten, & M.T. McCrudden (red.), *Handbook of multiple source use* (p. 447-465). New York/Londen: Routledge.
- Sabatini, J., O'Reilly, T., Weeks, J., & Wang, Z. (2020). Engineering a twenty-first century reading comprehension assessment system utilizing scenario-based assessment techniques. *International Journal of Testing*, 20(1), p. 1-23.
- Salmerón, L., & Llorens, A. (2019). Instruction of digital reading strategies based on eye-movements modeling examples. *Journal of Educational Computing Research*, 57(2), 343-359.
- Salmerón, L., Delgado, P., & Mason, L. (2020). Using eye-movement modelling examples to improve critical reading of multiple webpages on a conflicting topic. *Journal of Computer Assisted Learning*, 36(6), 1038-1051.
- Salmerón, L., Vidal-Abarca, E., Martínez, T., Mañá, A., Gil, L., & Naumann, J. (2015). Strategic decisions in task-oriented reading. *The Spanish Journal of Psychology*, 18, 1-10.
- Sanford, A.J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, 6(9), 382-386.
- Saudargas, R. (1997). *State-Event Classroom Observation System (SECOS). Observation manual*. Knoxville: University of Tennessee.
- Schaffner, E., & Schiefele, U. (2013). The prediction of reading comprehension by cognitive and motivational factors: Does text accessibility during comprehension testing make a difference?. *Learning and Individual Differences*, 26, 42-54.
- Schoot, M. van der, Reijntjes, A., & Lieshout, E.C. van. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*, 25(7), 1665-1690.
- Schotter, E.R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once); comprehension is supported by regressions during reading. *Psychological Science*, 25(6), 1218-1226.
- Schram, D. (red.). (2007). *Lezen in het VMBO*. Delft: Eburon.
- Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology*, 103(4), 877-896.
- Schunk, D.H. (1987). Peer models and children's behavioral change. *Review of Educational Research*, 57(2), 149-174.
- Schunk, D.H., & Zimmerman, B.J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading & Writing Quarterly*, 23(1), 7-25.
- Schuppert, L., & Bergh, H. van den. (2017). Leesstrategieën: Evidence-based of Nieuw Nederlands? Een experimentele studie in 4 vwo. *Levende Talen Tijdschrift*, 18(1), 14-23.
- Schuurs, U. (2019). De constructie van een (alternatieve) leestoets: verslag en opbrengsten. (Intern verslag.) Arnhem: Cito.
- Schuurs, U., & Breij, B. (2013). Retorische analyse. In J. Karreman & R. van Enschoot (red.), *Tekstanalyse: Methoden en toepassingen* (p. 99-150). Assen: Koninklijke Van Gorkum.
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Mifflin.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* 74, 107-120. doi: 10.1007/s11336-008-9101-0.

- Silfhout, G. van. (2014). *Fun to read or easy to understand? Establishing effective text features for educational texts on the basis of processing and comprehension research*. Dissertatie. Utrecht: LOT.
- Silfhout, G. van. (In voorbereiding.) Toetsing van taalonderwijs; toetsing als kans voor het leren en ontwikkelen in taal. In J. Dera, J. Gubbels, J. van der Loo, & J. van Rijt (red.), *Vaardig met vakinhoud: handboek vakdidactiek Nederlands*. Bussum: Coutinho.
- Silfhout, G. van, Evers-Vermeul, J., & Sanders, T. (2015). Connectives as processing signals: How students benefit in processing narrative and expository texts. *Discourse Processes*, 52(1), 47-76.
- Slagter, P. (2001). *Learning by instructing: studies in instructed foreign language teaching*. Rodopi: Amsterdam.
- SLO. (2016a). Position paper SLO t.b.v. hoorzitting/rondetafelgesprek examinering van het vak Nederlands in het voortgezet onderwijs d.d. 10 november 2016. Kamerstuk Tweede Kamer der Staten Generaal. Geraadpleegd via [tweedekamer.nl/kamerstukken/detail?id=2016Z20696&did=2016D42622](https://www.tweedekamer.nl/kamerstukken/detail?id=2016Z20696&did=2016D42622)
- SLO. (2016b). *Karakteristieken en kerndoelen onderbouw voortgezet onderwijs*. Enschede: SLO.
- SLO. (2017). *Leerplankundige analyse PISA 2015*. Enschede: SLO.
- SLO. (2021a). *Leesvaardigheid onderbouw vo; Domeinbeschrijving ten behoeve van peilingsonderzoek*. Amersfoort: SLO.
- SLO. (2021b). *Startnotitie Nederlands bovenbouw voortgezet onderwijs*. Amersfoort: SLO.
- Smeets, E. (2021). *Monitor hybride onderwijs: rapportage over het voortgezet onderwijs*. Nijmegen: KBA.
- Sosnoski, J. (1999). Hyper-readers and their reading engines. In G. Hawisher & C. Selfe (red.), *Passions, Pedagogies, and 21st Century Technologies* (p. 161-177). Utah: Logan.
- Steenbakkers, J., & Glopper, K. de. (2020). Vakdidactiek, pedagogiek en het schoolvak Nederlands: Persoonsvorming? #hoedan? (deel 2). *Levende Talen Magazine*, 107(2), 5-9.
- Steenbakkers, J., Stukker, N., & Glopper, K. de. (2020). Formuleerproblemen in leerlingewerk en schoolboeken Nederlands: Een onderzoek naar de aansluiting tussen de leerstof en de formuleerproblemen van leerlingen. *Pedagogische Studiën*, 98, 27-45.
- Stensel, R. van, & Houtveen, T. (2020). De vele kanten van leesbegrip; Verslag van een literatuurstudie naar kernelementen van effectief onderwijs in begrijpend lezen. *Levende Talen Tijdschrift*, 21(3), 3-12.
- Stellwag, H.W.F. (1955). *Selectie en selectiemethoden: een inleidende studie in het aansluitingsvraagstuk L.O. en V.H.M.O.* Groningen: J.B. Wolters.
- Stemmers, F.G.A. (1938). De leerschool der verving. *Levende Talen*, 99-109.
- Stevens, N. (2018). *Grip op tekstbegrip: Een onderzoek naar het tekstbegrip van proza en poëzie en de ontwikkeling ervan onder vwo-leerlingen in leerjaar 3 en 5*. Masterscriptie. Utrecht: Universiteit Utrecht.
- Talent. (2017). *Talent: vakboek bovenbouw Nederlands: havo-vwo-gym*. 's-Hertogenbosch: Malmberg.
- Traub, R.E. (1993). On the equivalence of the traits assessed by multiple-choice and constructed-response tests. In R.E. Bennet & W.C. Ward (red.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (p. 29-44). Hillsdale, NJ: Erlbaum.
- Trizano-Hermosilla, I., & Alvarado, J.M. (2016). Best alternatives to Cronbach's alpha re-

- liability in realistic conditions: congeneric and asymmetrical measurements. *Frontiers in Psychology*, 7, 769.
- Truijens, A. (2020, 28 september). We weten wat moet gebeuren om de leesramp te keren. Nu nog bedenken hoe. *de Volkskrant*.
- Udo, G.J., Bagchi, K.K., & Kirs, P.J. (2011). Using SERVQUAL to assess the quality of e-learning experience. *Computers in Human Behavior*, 27(3), 1272-1283.
- Vakdidactiek Geesteswetenschappen. (2022). *Dudoc Alfa*. Geraadpleegd op 27 juni 2022 via dudoc-alfa.vakdidactiegw.nl/
- Ven, P.H. van de. (1996). *Moedertaalonderwijs: interpretaties in retoriek en praktijk, heden en verleden, binnen- en buitenland*. Groningen: Wolters-Noordhoff.
- Vermeij, L. (2011). Het verbeteren van tekstbegrip met behulp van sorteertaken. *Levende Talen Tijdschrift*, 12(2), 33-41.
- Vernooy, K. (2022). *Begrijpend lezen op de basisschool. De tekst centraal*. Antwerpen/'s-Hertogenbosch: Gompel&Scavina.
- Vidal-Abarca, E., Mañá, A., & Gil, L. (2010). Individual differences for self-regulating task-oriented reading activities. *Journal of Educational Psychology*, 102(4), 1-10. doi: 10.1037/a0020062.
- Vidal-Abarca, E., Martinez, T., Salmerón, L., Cerdán, R., Gilabert, R., Gil, L., Mánia, A., Llorens, A.C., & Ferris, R. (2011). Recording online processes in task-oriented reading with Read&Answer. *Behavior Research Methods*, 43(1), 179-192.
- Walczyk, J.J., & Griffith-Ross, D.A. (2007). How important is reading skill fluency for comprehension? *The Reading Teacher*, 60(6), 560-569.
- Wallot, S., O'Brien, B.A., Haussmann, A., Kloos, H., & Lyby, M.S. (2014). The role of reading time complexity and reading speed in text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(6), 17-45.
- Warps, J., Visser, M. de, Lodewick, J., & Termorshuizen, T. (2021). *Verkenning vervolgonderwijs*. Nijmegen: ResearchNed.
- Welgraven, E. & Vos, B. de. (2019, 27 augustus). Laat leerlingen actief met een tekst bezig zijn, in plaats van die eindeloze vragen. *Trouw*.
- Wesdorp, H. (1981). *Evaluatietechnieken voor het moedertaalonderwijs: een inventarisatie van beoordelingsmethoden voor de stelvaardigheid, het begrijpend lezen, de spreek-, luister- en discussievaardigheid*. 's-Gravenhage: Staatsuitgeverij.
- Westhoff, G. (2012). Mesten en meten in leesvaardigheidstraining: Leesvaardigheidsonderwijs en examentraining zijn twee verschillende dingen. *Levende Talen Magazine*, 99(4), 16-20.
- Wetering, T. van de, & Groenendijk, R. (2015). Van toetsgerichte lezers naar zelfstandige, analytische lezers: Geen trucje maar een vaardigheid. *Levende Talen Magazine*, 102(2), 4-8.
- Wijekumar, K.K., Beerwinkle, A.L., Harris, K.R., & Graham, S. (2019). Etiology of teacher knowledge and instructional skills for literacy at the upper elementary grades. *Annals of Dyslexia*, 69(1), 5-20.
- WHO [World Health Organisation]. (2022). Infodemic. Geraadpleegd op 22 juni 2022 via who.int/health-topics/infodemic#tab=tab_1.
- Wilkerson, I.A.G., & Son, E.H. (2011). A dialogical turn in research on learning and teaching to comprehend. In M.L. Kamil, P.D. Pearson, E.B. Moje, & P.P. Afflerbach (red.), *Handbook of reading research (Vol. 4, p. 359-387)*. New York: Routledge.
- Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics*, 14(1), 1-33.

- Witte, T. (2018). De kunst van het onderwijzen; Naar een levenskrachtige visie op het literatuuronderwijs. *Nederlandse Letterkunde*, 23(3), 359 – 383.
- Wolf, M. (2018). *Reader, come home: The reading brain in a digital world*. New York: Harper.
- Wolf, M., & Barzillai, M. (2009). The importance of deep reading. *Educational Leadership, Literacy 2.0*, 6(66), 32–37.
- Yeari, M., Broek, P. van den, & Oudega, M. (2015). Processing and memory of central versus peripheral information as a function of reading goals: Evidence from eye-movements. *Reading and Writing*, 28(8), 1071-1097.
- Zimmerman, B.J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology*, 94(4), 660.

BIJLAGEN

Inhoudsopgave

Bijlage 1: Stimuli eye-track-experiment (hoofdstuk 2 en 3)	225
Bijlage 2: Twee sets teksten met vragen (hoofdstuk 4)	231
Bijlage 3: Een voorbeeldtaak uit de experimentele lessenserie (hoofdstuk 6)	241
Bijlage 4: Schattingen voor stellingen afzonderlijk (hoofdstuk 6)	244
Bijlage 5: Fit en passingsvergelijking van de modellen (hoofdstuk 6)	246
Bijlage 6: Effect interventie op tekstbegrip grafisch weergegeven (hoofdstuk 6)	248

BIJLAGE 1

Stimuli eye-track-experiment (hoofdstuk 2 en 3)

Tekst 1 (niveau 2F) Complotdenkers zien de wereld anders

[1] Op onze aardbol lopen een hoop complotdenkers rond die de meest bizarre dingen beweren. Zo was de maanlanding in 1969 in scène gezet en zijn er plannen om Europa de komende decennia helemaal Duits te maken. Die theorieën zijn natuurlijk best leuk maar sommige gaan wel erg ver. Onderzoekers braken zich lange tijd het hoofd over wat sommige mensen wel complotdenkers maakt en anderen niet. Nu denkt een team psychologen van de Vrije Universiteit het antwoord te weten: complotdenkers gebruiken een bepaald denkpatroon – ‘illusoire patroonherkenning’ – overdreven veel.

[2] Het team kwam tot deze conclusie na het uitvoeren van verschillende experimenten op 264 volwassenen. Zo testten de psychologen of de deelnemers verzonnen complottheorieën geloofden op een schaal van 1 tot 9. Vervolgens onderzochten de wetenschappers of deelnemers sneller patronen zagen in willekeurige dingen, zoals een muntje opgooien. Mensen die bij het eerste onderzoek hoog scoorden, waren ook geneigd hierin patronen te zien. Deze bevindingen werden bevestigd in nog drie experimenten.

[3] De conclusie die het team uit deze experimenten trok, is dat complotdenkers te veel en te vaak verbanden zien tussen gebeurtenissen die er eigenlijk niet zijn: ‘illusoire patroonherkenning’. Iedereen heeft dat weleens. Je droomt bijvoorbeeld over je opa en de dag erna hoor je dat hij is overleden. Dit kan haast geen toeval zijn, denk je dan. Maar complotdenkers gaan in dit ontdekken van patronen te ver.

[4] Sociaal psycholoog Michiel van Elk merkt naar aanleiding van dit onderzoek op: ‘Psychologen onderzoeken al langer de denkwijze van complotdenkers. Veel wetenschappers hebben geen verband gevonden, in dit onderzoek hebben ze dat wel. De vraag is hoe sterk de aangetoonde relatie is. Toevalstreffers kun je immers niet uitsluiten. En andere verklarende factoren blijven volledig buiten beschouwing.’

Naar: Laurien Onderwater, Kijk, geplaatst op 24 oktober 2017

Vragen bij ‘Complotdenkers zien de wereld anders’

Vraag 1 <1F/2F>

‘Zo was de maanlanding in 1969 in scène gezet en zijn er plannen om Europa de komende decennia helemaal Duits te maken.’ Waarvan zijn deze twee uitspraken voorbeelden?

Vraag 2 <2F>

‘complotdenkers gebruiken een bepaald denkpatroon – “illusoire patroonherkenning” overdreven veel.’ (alinea 1) In welke alinea wordt kritiek op deze conclusie geleverd?

Vraag 3 <3F>

‘Het team kwam tot deze conclusie na het uitvoeren van verschillende experimenten op 264 volwassenen.’ Twee van deze deelonderzoeken worden beschreven in het artikel. Leg uit hoe deze twee deelonderzoeken **samen** de conclusie van de wetenschappers ondersteunen.

Geef antwoord in één zin.

Vraag 4 <3F>

De tekst beschrijft een wetenschappelijk onderzoek. Citeer de zin die de hoofdvraag bevat van de wetenschappers bij hun onderzoek.

Vraag 5 <4F>

‘De vraag is hoe sterk de aangetoonde relatie is. [...] Andere factoren blijven volledig buiten beschouwing.’ Leg uit welke **algemene** kritiek op de conclusie van het onderzoek hiermee wordt geleverd. Geef antwoord in één zin.

Correctiemodel bij ‘Complotdenkers zien de wereld anders’

Vraag 1

Van complottheorieën / van bizarre dingen die complotdenkers bedenken

Vraag 2

Alinea 4

Vraag 3

Mensen die complottheorieën geloofden, zagen sneller patronen in willekeurige dingen.

Vraag 4

‘Onderzoekers braken ... anderen niet’.

Vraag 5

Andere factoren kunnen complotdenken mogelijk nog veel sterker veroorzaken dan illusoire patroonherkenning. / Mogelijk veroorzaakt illusoire patroonherkenning enkel in lichte mate complotdenken. / Het onderzoek is niet betrouwbaar genoeg.

Tekst 2 (niveau 2F) Super origineel in je skinny jeans

Helaas heeft de auteur geen toestemming gegeven voor opname van onze bewerking van deze tekst in dit boek.

Naar: Nathalie Wouters, NRC Handelsblad, 8 september 2017

Vragen bij 'Super origineel in je skinny jeans'

Vraag 1 <2F>

Leg uit wat de schrijver met alinea 1 wil aantonen. Geef antwoord in één zin.

Vraag 2 <3F>

Leg uit waarom volgens de tekst 'het huidige meisjesuniform zo alledaags en casual is'. Geef antwoord in één (korte) zin.

Vraag 3 <3F>

Internationale ketens maken 'het juist makkelijker dan ooit om een uitgesproken outfit aan te schaffen'. Met welke eerdere vaststelling is deze zin, strikt gezien, in tegenspraak? Citeer de betreffende zin.

Vraag 4 <3F>

Leg uit waarom het 'steeds homogener wordende aanbod' volgens de tekst leidt tot 'homogeen kleedgedrag'. Geef antwoord in één zin.

Correctiemodel bij 'Super origineel in je skinny jeans'

Vraag 1

Het sterk homogene kleedgedrag onder jongeren/tienermeisjes

Vraag 2

De mode is democratischer geworden, waardoor mensen zich vrijer/informeler kleden. / We zijn vrijer geworden.

Vraag 3

'Dat homogene ... wordende aanbod.'

Vraag 4

Iedereen koopt zijn kleding bij dezelfde (kleine groep) winkels.

Niet goed: Het aanbod aan kleding is helemaal homogeen geworden.

Tekst 3 (niveau 3F) Alcohol maakt u lelijker dan u lief is

[1] Onlangs kwam de Gezondheidsraad met een alarmerend rapport. Dat er iets goed zou zijn aan onze 'lievelingsverslaving' is een groot misverstand. Bepaalde vormen van kanker, verkeersongelukken, agressiviteit – het kan echt allemaal aan alcohol gerelateerd worden.

[2] Moet de overheid nu een grootschalig ontmoedigingsbeleid invoeren via belasting en voorlichting? Let intussen op de valse dubbele moraal: wel waarschuwen en belasten met accijnzen, maar niet volledig verbieden, ondanks de hoge kosten voor de gezondheidszorg. Voor het actief ontmoedigen van ongeremde consumptie pleit veel, maar dat geldt bijvoorbeeld ook voor frisdrank of worst. De vraag is bovendien of het wel de moeite waard is om overheidsgeld in te zetten op het terugdringen van lichte alcoholconsumptie. De meest recente poging tot een volledig verbod, de *Prohibition* (1920-1933) in de VS, bleek onhaalbaar en ineffectief. Corruptie, geweld en criminaliteit namen sterk toe en de overheid kon de inkomsten niet missen.

[3] De *Prohibition* had vooral een morele inslag. Onze moraal heet nu gezondheid en uiterlijk. Nog steeds drinken acht op de tien Nederlanders. Of dat veel verandert door te waarschuwen voor de gezondheid ('Alcohol maakt u lelijker en slomer dan u lief is!'), is de vraag.

[4] Tijdens de *Prohibition* waren er voorstanders van *temperance*, matigheid, en ook nu gaat het om hoeveelheid. Mensen doen voortdurend dingen die risico's inhouden. Ze eten ongezond, bewegen te weinig, ondernemen gevaarlijke sporten. Het zit in de menselijke aard risico's te nemen en die eigenschap heeft ons ook veel vernieuwing gebracht. Misschien kan de overheid misschien wel beter volstaan met een waarschuwingssticker op de fles: drink niet meer dan een glas per dag.

Naar: Louise O. Fresco, NRC Handelsblad, 4 oktober 2017

Vragen bij 'Alcohol maakt u lelijker dan u lief is'

Vraag 1 <2F >

Vanuit welke vaststelling zou de overheid 'nu' wat alcohol betreft 'een grootschalig ontmoedigingsbeleid' moeten overwegen? Citeer de betreffende zin.

Vraag 2 <3F >

'Let intussen op de valse dubbele moraal'. Leg uit wat er 'vals' is aan de 'dubbele moraal' van de overheid. Geef antwoord in één zin.

Vraag 3 <3F/4F >

'Voor het actief ontmoedigen van ongeremde consumptie pleit veel, *maar dat geldt bijvoorbeeld ook voor frisdrank of worst.*' Leg uit welke kritische opmerking de schrijfster met deze laatste toevoeging maakt bij het eventueel actief ontmoedigen van alcohol. Geef antwoord in één zin.

Vraag 4 <4F>

‘Moet de overheid nu een grootschalig ontmoedigingsbeleid invoeren via belasting en voorlichting?’

Geef aan welke twee algemeen-maatschappelijke rollen van de overheid indirect in de tekst tegenover elkaar worden gezet. Gebruik per maatschappelijke rol één zin.

Correctiemodel bij ‘Alcohol maakt u lelijker dan u lief is’

Vraag 1

‘Dat er ... groot misverstand.’ Of: ‘Bepaalde vormen ... gerelateerd worden.’

Vraag 2

Het is vals om aan iets ongezonds geld te verdienen / Het is vals om alleen te waarschuwen als iets zó ongezond is

Vraag 3

Er is (als je hierom alcohol verbiedt) veel meer wat je dan zou moeten ontmoedigen.

Vraag 4

- Een overheid die haar burgers actief beschermt / die bij burgers verkeerd/ongezond gedrag sterk tegengaat

- Een overheid die niet ingrijpt in het gedrag van burgers / die haar burgers (keuze)vrijheid biedt / iets gunt / die hoogstens verkeerd gedrag ontmoedigt

Tekst 4 (niveau 4F) Genderneutraliteit

[1] Genderneutraliteit of -fluiditeit is al een flinke tijd in de mode. Catwalks stromen vol met androgynen modellen. Modeketens maken *ungendered* kledinglijnen. Er is meer ruimte dan ooit om te experimenteren met je mannelijkheid of vrouwelijkheid en alles wat daar tussenin zit. En dat is in principe een uitstekende ontwikkeling. Transmensen krijgen het dan wat makkelijker in deze wereld: ze voelen zich vaak onprettig in hun lichaam en een *coming out*-proces is sociaal en soms ook medisch gezien nogal ingrijpend.

[2] Maar er zitten rauwe randjes aan deze ontwikkeling. Nu lijkt de groep die ook zonder hype al transseksueel of nonbinair was, een minderheid aan het worden binnen dat alom aanwezige frivole geëxperimenteer met geslacht. Een nieuwe generatie wil niet meer in de man-vrouw-hokjes passen, maar meet zichzelf een identiteit op maat aan. Met een geslacht op maat, seksualiteit op maat en in veel gevallen een houding waarbij iedereen die zich vergist in een van de vijftig nieuwe genderidentiteiten of worstelt met de nieuwe taal een soort misdaad tegen de menselijkheid begaat. Bij gebrek aan niet-blanke huidskleur of non-conformistische religie kaapt dit nogal boze, nogal extreme groepje *social justice*-activisten de genderstrijd.

[3] Hún strijd, die gepaard gaat met scherp taalgebruik en nieuwe sociale omgangsvormen, verdiept de kloof binnen onze samenlevingen. Nu dreigt het ene hokje te worden afgebroken, terwijl het andere weer net zo hard wordt opgebouwd. Zonde.

Vragen bij 'Genderneutraliteit'

Vraag 1 <2F>

Welke twee voorbeelden geeft de tekst van het feit dat genderneutraliteit al 'een flinke tijd in de mode' is? Citeer de betreffende zinnen.

Vraag 2 <2F>

De recente aandacht voor genderneutraliteit is volgens de tekst 'in principe een uitstekende ontwikkeling'. Citeer de zin die aangeeft waarom het een uitstekende ontwikkeling is.

Vraag 3 <3F>

De 'genderstrijd' wordt 'gekaapt'. Van wie wordt deze genderstrijd weggekaapt? Citeer een woordgroep.

Vraag 4 <3F/4F>

Naast problemen voor groepen in de samenleving, veroorzaakt de ontwikkeling rondom genderneutraliteit volgens de tekst een mogelijk probleem voor **iedereen**. Citeer de zin die dit algemeen-maatschappelijke probleem benoemt.

Vraag 5 <4F>

'Nu dreigt het ene hokje te worden afgebroken'. Leg deze beeldspraak uit en geef daarbij aan wiens 'hokje' wordt afgebroken. Geef antwoord in één zin.

Correctiemodel bij 'Genderneutraliteit'

Vraag 1

"Catwalks stromen vol met androgynе modellen. Modeketens maken *ungendered* kledinglijnen."

Vraag 2

'Transmensen krijgen ... deze wereld.'

Vraag 3

De transmensen / de groep die ook zonder hype transseksueel of nonbinair was

Vraag 4

'Hun strijd ... onze samenlevingen.'

Vraag 5

Het isolement van transmensen wordt opgeheven / Transmensen worden geaccepteerd / wijken niet meer af. / Het (binaire) man-vrouwdenken wordt losgelaten.

BIJLAGE 2

Twee sets teksten met vragen (hoofdstuk 4)

Set 1: afkomstig uit Centraal Examen Nederlands vwo 2019-1; bewerkt

De kunst en de regels

[1] Vorige week woonde ik een opvoering bij van *Mariken van Nimwegen*¹, want mijn zoon was er figurant (een heel goede, concludeerden we achteraf). Zo leerde ik in de huiselijke Stadsschouwburg van Groningen het verhaal van Mariken kennen. Een lacune in mijn academische opleiding die tegenwoordig niet meer mogelijk is: het internet wemelt van de leesdossiers waarin scholieren je een samenvatting plus exegese geven. Het is maar dat u het weet, mocht u ook iets in te halen hebben.

[2] De kernregel uit Marikens gedicht 'Door d'onkonstige gaat die konste verloren'² is een noodkreet van begin zestiende eeuw die vreemd eigentijds aandoet. Mariken heeft het over de 'seven vrie consten: rethorijcke, musijcke, logica, gramatica ende geometrie, arismatica ende alkenie'³. Ik zou daar vijf eeuwen later de kunst van goed onderwijs, onderzoek, zorg en zo meer aan willen toevoegen. Wie daarvan de kunst niet verstaat, kan volgens Mariken slechts schade aanrichten. Het is dus een oude, maar helaas brandende kwestie.

[3] Een paar voorbeelden uit de zorg. Een huisarts meldt dat hij vijf A4'tjes moest invullen om dieetvoeding voor een terminale patiënt te bemachtigen. De minutenregistratie van zorg 'achter de voordeur geleverd' is berucht. Hoe moet de tijd van de ingang van een flat naar de patiënt op de tiende etage geregistreerd worden – wat geldt hier als voordeur? Of het consult met de huisarts op diens kantoor, wat dus geen achter de voordeur geleverde zorg is? De regels van hen die de kunst niet verstaan, staan hier de kunst van goede zorg in de weg. Het nare is dat de regels het snelst groeien waar ze de meeste schade aanrichten. Waarom?

[4] Noem het de Wet van Mariken. Ze draait om het begrip *tacit knowledge*⁴, een idee van Michael Polanyi. Het gaat om ervaringskennis die niet in een boekje is op te schrijven en niet overdraagbaar is door een paar uur instructie of inwerken. Vakken vullen, bollen pellen en boompjes stekken vergen vrijwel geen ervaringskennis (ik kom uit de buurt van Boskoop, dus spreek uit ervaring). Je kunt vanaf dag één aan de slag. Voor meesters en juffen, wetenschappers en verzorgenden is dat wel anders – en de lijst is natuurlijk veel langer. Want wat is een goede docent, verzorgende, of onderzoeker? Dat is lastig te zeggen – en dus is er volop ruimte voor regels die dat specificeren. (Denkt u dat zulke regels er ook zijn voor onkruidwieders en vakkenvullers? Welnee. Daar zie je direct of het werk goed gedaan is.)

[5] Tegelijk zijn regels het schadelijkst juist daar waar veel ervaringskennis is. Want de regels van onwetende buitenstaanders dwingen de ervaringsdeskundigen hun werk naar

de verkeerde maatstaven te beoordelen en uit te voeren, en daar bovendien veel tijd aan te besteden. Denk aan het construeren van ranglijstcriteria, het tellen van stappen en minuten, en het aanvinken van vakjes in grote tabellen (ik spreek opnieuw uit ervaring). Meestal zonde van de tijd, en geen kunst. Maar er is meer.

[6] Weet u wie daar het slechtst tegen kunnen? Juist, de meest ‘konstigen’. De mensen die om de leerlingen het onderwijs in gingen, die uit nieuwsgierigheid wetenschapper werden, en die in de zorg werken omdat ze willen verzorgen. Dat zijn de mensen die de moed het eerst in de schoenen zakt. Ziedaar de andere kant van de Wet van Mariken: regels groeien niet alleen sneller waar meer ervaringskennis aan het werk is, maar ze jagen er ook juist de mensen met passie en toewijding weg.

[7] Dat is geen toeval. Het opbouwen van ervaringskennis vraagt intrinsieke motivatie. Wie intrinsiek gemotiveerd is, ontleent bevrediging aan werk dat goed gedaan wordt. En wordt dus sneller gedemotiveerd door regels die het verkeerde eisen en meten. Die persoon kan het niet opbrengen nóg eens de les aan te passen aan de Citotoets, de vergoedingscriteria van de verzekeraar na te kijken, of de impactscore van een wetenschappelijk tijdschrift te vermenigvuldigen met het aantal publicaties. Die gaat dan maar iets anders doen.

[8] Maakt u zich geen zorgen om de anderen, de mensen die hun werk vooral als bron van inkomen zien. Zij redden het wel. Maar wees eerlijk: dit zijn de ‘onkonstigen’. Waar zij de regels gaan bepalen – en, als het even kan en met de beste bedoelingen, ook uitbreiden, ‘optimaliseren’ en in ‘targets’ vertalen – daar wordt het slecht toeven. Daar gaat de kunst verloren, leert ons de Wet van Mariken.

Naar: Dirk Bezemer. Uit: de Groene Amsterdammer, 25 november 2015

noot 1 Mariken van Nimwegen is een bekend Nederlands zestiende-eeuwse mirakelspel, een middel-eeuwse vorm van toneel waarin een heiligenleven wordt verbeeld of waarin een wonder centraal staat.

noot 2 ‘Door d’onkonstige gaat die konste verloren’ betekent vrij vertaald: Door de onkundigen gaat de kunst verloren.

noot 3 ‘seven vrie consten: rethorijcke, musijcke, logica, gramatica ende geometrie, arismatica ende alkenie’ betekent: de zeven vrije kunsten: retorica, muziek, logica, grammatica en meetkunde, rekenkunde en alchemie.

noot 4 tacit knowledge: kennis die alleen dankzij ervaring kan ontstaan en vaak intuïtief wordt toegepast.

Vragen bij De kunst en de regels

In alinea 4 van de tekst wordt het begrip *tacit knowledge* geïntroduceerd.

- 2p **1** Welke vier van onderstaande groepen zouden vooral beschikken over *tacit knowledge*, gelet op de strekking van de tekst? Kies uit: bollenpellers, buitenstaanders, ‘konstigen’, managers, onderwijzers, ‘onkonstigen’, patiënten, scholieren, verzekeraars, verzorgenden en wetenschappers.

De titel van de tekst luidt ‘De Kunst en de Regels’.

- 2p **2** Leg uit hoe de Kunst zich verhoudt tot Regels, gelet op de strekking van de tekst. Geef antwoord in één zin.

- 1p 3 Wat is het tekstdoel van de tekst, gelet op de strekking van de tekst?
 'De Kunst en de Regels' is een
- A beschouwing, want de kernregel uit Mariken van Nimwegen wordt vanuit verschillende invalshoeken in een moderne context geplaatst.
 - B beschouwing, want van diverse beroepen in onderwijs, onderzoek en zorg wordt bekeken of de Wet van Mariken erop van toepassing is.
 - C betoog, want aan de hand van de Wet van Mariken wordt gepropageerd dat het snel groeiende aantal regels bij ervaringsberoepen een kwalijke zaak is.
 - D betoog, want door het geven van schrijnende voorbeelden wordt bepleit dat we werk niet vooral als bron van inkomen moeten zien.

In de tekst komen praktijkgerichte beroepen en beroepen met veel *tacit knowledge* ter sprake. Tussen beide beroepssoorten staan enkele inhoudelijke tegenstellingen centraal.

- 3p 4 Noem drie van deze inhoudelijke tegenstellingen. Vul dit in in onderstaand schema.

praktijkgerichte beroepen	beroepen met veel <i>tacit knowledge</i>
1a)	1b)
2a)	2b)
3a)	3b)

In de tekst wordt de Wet van Mariken besproken. Deze bestaat volgens de tekst uit twee wetmatigheden.

- 2p 5 Benoem elk van deze wetmatigheden.
 Geef antwoord in een of meer volledige zinnen.
- 2p 6 Leg uit in welke twee opzichten de door 'onkonstigen' opgestelde regels het werk voor 'konstigen' in de praktijk moeilijk maken.
 Geef geen voorbeelden. Geef antwoord in een of meer volledige zinnen.

In alinea 4 tot en met 8 van de tekst wordt de houding besproken die de 'konstigen' en de 'onkonstigen' innemen ten aanzien van werk dat *tacit knowledge* vereist.

- 4p 7 Benoem in onderstaand schema wat volgens tekst 1 de essentiële verschillen in houding zijn tussen de leden van de twee groepen ten aanzien van werk dat *tacit knowledge* vereist.
 Geef antwoord in een of meer volledige zinnen.

de 'konstigen'	de 'onkonstigen'
1a	1b
2a	2b

- 2p 8 Wat wordt er, gezien de strekking van de tekst, in deze tekst bepleit?
- A dat de beoordeling van werk alleen wordt uitgevoerd door mensen die dat werk zelf doen of gedaan hebben
 - B dat 'konstigen' meer tijd aan hun werk kunnen besteden en niet aan irrelevante regels worden blootgesteld
 - C dat voor alle beroepen dezelfde regels en dezelfde beoordelingsnormen gaan gelden
 - D dat werk waarvoor veel ervaringskennis nodig is, wordt beoordeeld via uitgebreide, door vakgenoten opgestelde regels

(1) Jos de Blok, directeur van Buurtzorg, houdt vast aan het principe van kleinschaligheid. Veel thuiszorgondernemingen gaan volgens hem onderuit door een topzwaar apparaat, dat vooral gericht is op het controleren van het personeel. (...)

(2) "Mensen die ver afstaan van de dagelijkse praktijk, bedenken oplossingen, altijd uitgedrukt in geld, die los staan van de werkelijke problemen. De budgetten staan centraal, maar de vraag naar de beste weg voor de betrokken persoon komt nauwelijks aan bod. (...)

(3) Vanaf de jaren tachtig heeft in veel landen de gedachte postgevat dat je door bedrijfsmatig handelen komt tot betere oplossingen, meer kwaliteit en lagere kosten - wat je dan door allerlei marktprikkels afdwingt. Maar in al die landen zie je het omgekeerde gebeuren. De kosten zijn gestegen en er heerst ontevredenheid bij cliënten en professionals."

naar: Stevo Akkerman. Uit: Trouw, 1 maart 2016

In alinea 3 van tekstfragment 1 worden twee nadelen van de huidige praktijk genoemd die niet in de hoofdttekst aan de orde komen.

- 2p 9 Welke nadelen zijn dat?

Correctiemodel bij De Kunst en de regels

1 maximumscore 2

'konstigen', onderwijzers, verzorgenden en wetenschappers

indien vier goed 2p

indien drie goed 1p

indien twee of minder goed op

2 maximumscore 2

De kern van een goed antwoord is:

De Kunst wordt gehinderd door de Regels / Vaak hebben regels een negatief effect op werkzaamheden waar ervaringskennis voor nodig is / Hoe meer regels, des te eerder raakt kunst in het gedrang

3 C (1p)

4 maximumscore 3

drie van onderstaande inhoudelijke tegenstellingen:

praktijkgerichte beroepen

a) geen/nauwelijks regels

a) vrijwel geen ervaringskennis /
vrijwel geen tacit knowledge

a) gemakkelijk aan te leren

a) eenvoudig beoordeelbaar

a) geen schade als gevolg van regels

a) (kennis is) gemakkelijk overdraagbaar
per juist benoemde tegenstelling 1p

ervaringsgerichte beroepen

b) veel regels

b) veel ervaringskennis /
veel tacit knowledge

b) lang leertraject noodzakelijk

b) lastig beoordeelbaar

b) negatieve effecten als gevolg van regels

b) (kennis is) lastig overdraagbaar

5 maximumscore 2

De kern van een goed antwoord is:

a. Wetmatigheid 1 zegt dat regels het snelst groeien waar ze de meeste schade aanrichten / bij ervaringsgerichte beroepen (1p)

b. Wetmatigheid 2 zegt dat regels het schadelijkst zijn daar waar veel ervaringskennis nodig is (1p)

6 maximumscore 2

De kern van een goed antwoord is:

a) Deze regels kosten 'konstigen' veel tijd / werk (1p)

b) Deze regels meten het verkeerde / hebben geen meerwaarde (1p)

7 maximumscore 4

De kern van een goed antwoord is:

1a Zij doen hun werk met passie (en toewijding) / Zij zijn intrinsiek gemotiveerd.

1b Zij werken voor het geld / Zij missen de intrinsieke motivatie.

2a Zij raken gedemotiveerd door de bemoeienis/regels van onkundigen.

2b Zij hebben niet zo veel moeite met regels (en haken daardoor minder snel af).

Als 1a en 1b beide juist zijn: 2 scorepunten toekennen.

Als 2a en 2b beide juist zijn: 2 scorepunten toekennen. Als alleen 1a of 1b goed is: 1 scorepunt toekennen.

Als alleen 2a of 2b goed is: 1 scorepunt toekennen.

De koppeling van 1a en 2b levert 1 scorepunt op; ook de koppeling van 2a en 1b levert 1 scorepunt op.

8 B (2p)

9 maximumscore 2

De kern van een goed antwoord is:

a. de stijgende kosten (die de werkwijze met zich meebrengt) (1p)

b. de ontevredenheid bij cliënten (1p)

Let op: Indien 'en professionals' is toegevoegd aan het antwoord bij het tweede bolletje, dient dat antwoord fout te worden gerekend.

Set 2: afkomstig uit het Centraal Examen Nederlands havo 2017-2; bewerkt*De wilskrachtige is een gewoontedier*

[1] Wat eet je liever: een appel of een chocoladereep? Worteltjes of patat? De appel en de worteltjes zijn natuurlijk de gezondste keuze: caloriearm, vol vezels en vitamines. Zoete troep en vette hap zijn nergens goed voor, maar wél lekker. Dus wat kies je? Utrechtse psychologen legden hun proefpersonen onlangs keuzes als deze voor. En wat bleek?

[2] Mensen die van nature wilskrachtig zijn, kiezen vaker voor de appel en de worteltjes dan de minder wilskrachtigen, de uitslapers en de flapuits. Ze kiezen niet alleen vaker, maar ook sneller voor het verstandige alternatief. Ze twijfelen minder, voeren slechts een korte of helemaal geen innerlijke strijd.

[3] Dit nieuwe onderzoek sluit aan bij eerdere bevindingen van het *Selfregulation Lab*, zoals een groep psychologen aan de Universiteit Utrecht zichzelf noemt. Sinds 2002 bestuderen zij de wilskracht: het vermogen om impulsen te onderdrukken met het oog op een grotere beloning in de toekomst. Wie slank wil blijven, bezwijkt niet voor patat, gaat niet (te vaak) de kroeg in. Naast intelligentie is zelfcontrole een belangrijke sleutel tot zowel persoonlijk als maatschappelijk succes.

[4] Onder leiding van hoogleraar gezondheidspsychologie Denise de Ridder deed deze groep enkele jaren geleden al een interessante ontdekking: het kost wilskrachtige mensen vaak heel weinig moeite verleidingen het hoofd te bieden. Ze zijn geen mentale krachtpatsers die hun verlangens naar junkfood, alcohol en overspelige seks steeds met geweld weten te onderdrukken: ze vóelen die verlangens eenvoudigweg niet of veel minder dan hun zwakkere broeders en zusters. Innerlijke conflicten weten ze te omzeilen met routines.

[5] Daar hebben wilskrachtige mensen er namelijk relatief veel van, blijkt uit studies van het Utrechtse team. Ze hebben meer gezonde en minder ongezonde gewoontes dan de patateters en de kroegtijgers. Voor het ontbijt nemen ze yoghurt met muesli, rond koffietijd op kantoor een appel en geen gevulde koek. Dankzij de automatische piloot hoeven ze hun zelfbeheersing niet in stelling te brengen, zegt De Ridder. Hun tankje wilskracht blijft gevuld voor noodgevallen.

[6] Uit onderzoek blijkt dat dit tankje bij iedereen leeg kan raken. Vergelijk wilskracht met een spier: als je die flink belast, wordt hij moe. Ook goedgetrainde bodybuilders zijn weleens uitgeput, en moeten dan rusten om weer op krachten te komen.

[7] Het beeld van de spier is van Roy Baumeister, hoogleraar sociale psychologie aan de Vrije Universiteit in Amsterdam. Hij is ook de uitvinder van de term 'egodepletie': een verminderd vermogen om je gedrag en gevoelens te controleren. Bij egodepletie ben je (tijdelijk) door je voorraad wilskracht heen en voel je intensere verlangens naar verboden vruchten.

[8] In zijn boek *Wilskracht. De herontdekking van de grootste kracht van de mens* (2012) beschrijft Baumeister hoe hij egodepletie wist vast te stellen. Dat begon met het radijsexperiment: hij liet studenten eerst vasten, zodat ze flink honger hadden. Vervolgens nodigde hij ze uit in zijn psychologisch laboratorium, waar de geur hing van versgebakken chocoladekoekjes. Op tafel stonden drie schalen: één met de nog warme koekjes, één met chocolaatjes en één met radijsjes. Sommige studenten mochten aanvallen op de zoetigheden, anderen werd gevraagd zich tot de radijsjes te beperken.

[9] Psychologisch onderzoek is soms licht sadistisch, merkt Baumeister zelf op: hij liet de studenten met de radijso opdracht een tijdje alleen in het laboratorium en keek vervolgens door een verborgen raampje hoe zij reageerden op de koekjes. Ze bleken allemaal

bikkels: weliswaar keken ze hunkerend naar de koekjes, ze roken er soms aan, maar niemand nam er een hap van.

[10] Maar nu komt het: daarna moesten alle studenten in een andere ruimte meetkundeopgaven maken die onoplosbaar waren. Zelf dachten ze dat het om een intelligentietest ging. De studenten die hun maag met koekjes en chocola hadden gevuld, werkten zo'n twintig minuten aan de opdrachten, net als een controlegroep van hongerige studenten die helemaal niets te eten hadden gehad. 'De zwaarbeproefted radijseters gaven het echter al na acht minuten op - voor een laboratoriumexperiment een enorm verschil', schrijft Baumeister. 'Ze hadden weerstand weten te bieden aan de verleidelijke koekjes en chocolade, maar hadden door die inspanning minder energie over voor de meetkundeopgaven.'

[11] Uit dit en ander onderzoek leidden Baumeister en collega-psychologen verschillende adviezen af voor wie zijn wilskracht wil versterken, zoals: kies één doel tegelijk. Met een lege maag versgebakken koekjes weerstaan is nog wel te doen, maar daarna ook nog eens je hersens langdurig laten kraken, nee. Tip twee is het trainen van je wilskracht, want dat kán, zoals je ook een spier kunt opbouwen. Je kunt bijvoorbeeld elke dag een op zichzelf vervelend taakje doen: al je uitgaven bijhouden in een kasboek, twee weken lang rechtop zitten. Ze vormen allemaal een goede aanloop naar een grotere uitdaging, aldus Baumeister, 'zoals stoppen met roken of je houden aan een budget.'

[12] Hieraan voegt Denise de Ridder, die in Utrecht de routine ontdekte als geheim wapen van de wilskrachtige mens, nog een advies toe. 'Het lijkt er inderdaad op dat je wilskracht kunt trainen door dagelijks kleine, vervelende taakjes te doen. Maar ik denk dat het effectiever is routines op te bouwen waardoor je die wilskracht niet hoeft aan te spreken.' Ook al omdat de positieve effecten van wilskrachttoefeningen vaak maar kort duren; na verloop van tijd kan de zelfbeheersing weer verslappen. Daarom pleit De Ridder voor een nieuwe gewoonte: 'Neem elke keer als je zin hebt in een glas wijn, een glas van je favoriete vruchtensap. Je kunt de alcohol beter vervangen dan je verlangen ernaar onderdrukken, want dan blijf je denken: oei, lekker, glaasje wijn.'

Naar: Iris Pronk. Uit: Trouw, 4 juli 2015

Vragen bij De wilskrachtige is een gewoontedier

- In de openingsalinea's is sprake van recent Utrechts onderzoek.
- 2p 1 Welke nieuwe inzichten over wilskrachtige mensen gaf dit recente Utrechtse onderzoek?
Geef antwoord in een of meer volledige zinnen.
- In de eerste helft van de tekst wordt gesproken over recent én ouder onderzoek onder leiding van hoogleraar gezondheidspsychologie Denise de Ridder.
- 1p 2 Welke van de onderstaande omschrijvingen geeft de beste samenvatting weer van de resultaten van dat onderzoek?
Wilskrachtige mensen
- A hebben zoveel gezonde gewoontes dat ze de lokroep van ongezonde verleidingen minder horen en daardoor hun wilskracht nauwelijks hoeven aan te spreken.
- B onderdrukken hun verlangens naar ongezonde verleidingen door snel over te stappen op een automatische piloot die hen ervoor behoedt toe te geven aan die verleidingen.

- C voelen wel degelijk een bepaalde honger naar junkfood en alcohol, maar door een beroep te doen op hun voorraad wilskracht weten ze die drang om te zetten in gezonde gewoontes.
- D zijn vatbaarder voor verlangens dan mensen die bepaalde routines opgebouwd hebben omdat die laatste groep meer succeservaringen heeft geboekt op het gebied van zelfcontrole.

In de tekst valt de term 'egodepletie'.

- 1p 3 Welk van de onderstaande omschrijvingen geeft het best weer wat egodepletie te maken heeft met wilskracht?
- A Als je wilskracht niet voldoende gebruikt, krijg je een tekort aan egodepletie.
- B Als je je voorraad wilskracht hebt opgebruikt, kun je spreken van egodepletie.
- C Als je je wilskracht op de proef stelt, voorkom je een toestand van egodepletie.
- D Als je wilskracht regelmatig traint, kom je sneller in aanmerking voor egodepletie.

In de tekst worden drie algemene adviezen gegeven hoe je wilskrachtiger kunt worden.

- 3p 4 Geef aan wat die drie algemene adviezen zijn.
Geef antwoord in een of meer volledige zinnen.

In de eerste helft van de tekst wordt uitgelegd dat er voordelen zitten aan het worden van een gewoontedier.

- 3p 5 Noem de voordelen uit tekst 2.
Geef antwoord in een of meer volledige zinnen.

Sinds 2002 bestuderen zij de wilskracht: het vermogen om impulsen te onderdrukken met het oog op een grotere beloning in de toekomst.' (al. 3)

Deze definitie van wilskracht lijkt niet goed aan te sluiten op de beschrijving van wilskrachtige mensen elders in de tekst.

- 1p 6 Leg uit in welk opzicht de definitie van wilskracht niet aansluit op het in het onderzoek beschreven gedrag van wilskrachtige mensen.
Geef antwoord in een of meer volledige zinnen.

Tekstfragment 1

Mensen die obees zijn, worden vaak beschuldigd van het feit dat zij geen wilskracht en discipline zouden hebben. Zelfs sommige gezondheidsprofessionals en huisartsen zijn daar schuldig aan. Het is wat ons betreft onzin dat mensen met overgewicht minder wilskracht of discipline zouden hebben. Veel mensen met obesitas zijn bijvoorbeeld zakelijk zeer succesvol, maar laten de teugels op andere gebieden 'vieren'. Hebben zij dan minder wilskracht dan iemand die een toplijf en topconditie heeft, maar op andere gebieden misschien minder succes boekt? Nee, hoogstwaarschijnlijk niet.

naar: <http://resetpersonaltraining.nl>

In tekstfragment 1 wordt een probleem aan de orde gesteld dat ook in de hoofdtekst aan de orde kwam.

- 2p 7 Leg uit welke verklaring voor dit probleem uit tekstfragment 1 kan worden gegeven op grond van de tekst.
Geef antwoord in een of meer volledige zinnen.

De titel van de tekst is 'De wilskrachtige is een gewoontedier'.

- 1p 8 Wordt deze bewering in de tekst onderbouwd?
- A Ja, want er wordt aangetoond dat wilskrachtigen verleidingen vermijden met routines.
 - B Ja, want uit onderzoek blijkt dat zelfcontrole een belangrijke sleutel tot succes is.
 - C Nee, want de werking van de wilskracht wordt wel toegelicht, maar een voorbeeld ontbreekt.
 - D Nee, want er worden enkel voorbeelden gegeven van de werking van de wilskracht als spier.
- 1p 9 Welke van de onderstaande omschrijvingen typeert het doel van de tekst het best?
- De tekst heeft tot doel de lezer
- A ervan te overtuigen dat deze op basis van wilskracht ongezonde gewoontes moet omzetten in gezonde routines.
 - B op de hoogte te stellen van uitkomsten van recent onderzoek naar de invloed van wilskracht en routines op het leveren van prestaties.
 - C te laten inzien dat het hanteren van vaste gewoontes een efficiëntere manier is om gedrag aan te passen dan het trainen van wilskracht.
 - D te stimuleren tot het inzetten van wilskracht en tot het leven volgens vaste gebruiken, omdat dit succesfactoren zijn.

Correctiemodel bij De wilskrachtige is een gewoontedier

1 maximumscore 2

De kern van een goed antwoord is:

- Wilskrachtige mensen kiezen vaker het verstandige alternatief / beheersen zich vaker 1
- (en) doen dat sneller / twifelen veel korter over hun keuze / nemen sneller een besluit 1

2 A (1p)

3 B (1p)

4 maximumscore 3

De kern van een goed antwoord is:

- a. Kies één doel tegelijk (1p)
- b. Train je wilskracht (1p)
- c. Bouw routines op / Ontwikkel nieuwe gewoontes (1p)

5 maximumscore 3

De kern van een goed antwoord is:

- a) Het kost weinig moeite om verleidingen te weerstaan (1p)
- b) Ze kunnen innerlijke conflicten omzeilen (met routines) (1p)
- c) Hun wilskracht blijft beschikbaar voor noodgevallen / Hun tankje wilskracht blijft gevuld voor noodgevallen (1p)

6 maximumscore 1

De kern van een goed antwoord is:

Wilskrachtige mensen hoeven geen impulsen te onderdrukken, want ze voelen die niet.

7 maximumscore 2

De kern van een goed antwoord is:

- a) Doordat mensen al hun wilskracht in het bereiken van andere dingen stoppen (zoals zakelijke succes) / Doordat er veel wilskracht wordt gebruikt voor andere zaken (zoals zakelijk succes) (1p)
- b) lukt het niet meer om de wilskracht in te zetten (om af te vallen / om minder/gezonder te eten) / raakt de wilskracht op (om af te vallen / om minder/gezonder te eten) (1p)

Opmerking

Goed voor 2 scorepunten:

Er is sprake van egodepletie.

8 A 1

9 B 1

BIJLAGE 3

Een voorbeeldtaak uit de experimentele lessenserie

Opdracht 'Neanderthalers'

Doel van deze opdracht

Stel, je moet voor het vak geschiedenis **een presentatie geven over Neanderthalers** (je kent ze misschien wel: een uitgestorven menssoort) en je wilt onder meer weten of Neanderthalers konden praten. Om die vraag te beantwoorden, vind je op een wetenschapsinternetpagina een handige korte tekst die hierop een antwoord geeft. Je besluit deze tekst voor je presentatie te gebruiken.

Wat ga je daarvoor doen?

Bestudeer de tekst die je docent heeft uitgedeeld. **Bouw voor jezelf diep begrip van deze tekst** op: konden Neanderthalers nu waarschijnlijk praten, of toch niet? Je moet niet alleen deze vraag vanuit jezelf kunnen beantwoorden in je presentatie, maar ook **vervolgvragen** hierover vanuit het publiek na je presentatie.

Nadat je de tekst hebt bestudeerd, **haalt de docent de tekst op**. Je ontvangt dan enkele vragen, ook vragen die je over deze kwestie zou kunnen krijgen vanuit het publiek. *Beantwoord ze vanuit je opgebouwde begrip. De tekst mag en kun je dus niet meer inzien!*

Hebben Neanderthalers ooit met elkaar gepraat?

De kans dat Neanderthalers een vorm van taal hadden, is best wel groot. Mark Dingemans, universitair hoofddocent taal en communicatie aan de Radboud Universiteit en onderzoeker bij het Max Planck Instituut, denkt dat deze oermens in ieder geval klanken kon maken.

[1] Dat ze lange tijd werden gezien als stille wezens, kwam omdat Neanderthalers een heel kleine keelholte hebben. Te klein om klanken mee te produceren. Totdat er in 1989 een tongbeen van een Neanderthaler werd gevonden in de grot Kebara in Israël, waarmee de oermens in een klap een stem kreeg. 'Daaruit konden onderzoekers afleiden dat ze een spraakorgaan moeten hebben gehad dat lijkt op dat van de moderne mens. En dus konden ze spraak produceren', aldus Dingemans. 'Men heeft ook botten van de kaak en de schedel gevonden en een reconstructie gemaakt. Ook daaruit blijkt dat ze waarschijnlijk konden praten.'

[2] Toch plaatst Dingemane een kanttekening bij deze conclusies. Want klanken kunnen produceren, is nog niet hetzelfde als praten. ‘Hadden ze ook woorden? Maakten ze gebruik van gebaren? Luisterden ze naar elkaar en reageerden ze daarop? Die puzzelstukjes kunnen we niet leggen.’

[3] Volgens Dingemane is er wel indirect bewijs dat duidt op spraakvermogen bij de Neanderthaler. ‘Archeologen ontdekten bijvoorbeeld dat ze waarschijnlijk vuur onder controle konden houden. Dat is geen geringe prestatie, daar heb je best wat denkkracht voor nodig. En je moet ervoor samenwerken, dus er bestaat een kans dat ze daarvoor taal gebruikten. Het zijn interessante hypothesen en het klinkt plausibel, maar zeker weten doen we het niet.’

[4] Het zou makkelijker zijn als er zoiets bestond als een ‘taalgen’, maar dat is helaas niet het geval. ‘Het is niet zo dat een specifiek gen verantwoordelijk is voor ons taalvermogen, dus daar kunnen we niet naar zoeken bij Neanderthalers. We zijn er wel achter gekomen dat ze bepaalde versies van genen hadden die helpen bij het formuleren van klanken. Of ze ook daadwerkelijk konden praten blijft voorlopig giswerk.’ (...)

Naar: Roeliene Bos, Quest, 26 juni 2019

Opdrachten (de docent heeft de tekst nu opgehaald!)

1) Je geeft je presentatie. Wat ga je vertellen? In hoeverre staat het volgens jou vast dat Neanderthalers konden praten?

- A Nog onzeker
- B Zeer waarschijnlijk
- C Zeer onwaarschijnlijk
- D Zo goed als zeker

2) Wat is de belangrijkste reden dat het nog onzeker is of Neanderthalers konden praten?

- A De keelholte van Neanderthalers maakt praten erg moeilijk.
- B Het spraakorgaan van Neanderthalers lijkt niet op dat van de moderne mens.
- C Klanken kunnen maken is niet hetzelfde als praten.
- D Neanderthalers maakten vooral veel gebruik van gebaren.

Je hebt je presentatie gegeven en het liep lekker! (Hoera! Applaus ...) ... maar je publiek wil nog wat zaken van je weten ...

3) ‘Ik dacht dat Neanderthalers fysiek gewoonweg niet konden praten. Wat hebben ze dan ontdekt waardoor je denkt dat Neanderthalers toch konden praten?’

4) ‘Hangt spraakvermogen bij Neanderthalers volgens jou samen met genetische aanleg?’

5) ‘Ik heb weleens gelezen dat je uit het onderlinge gedrag kunt opmaken dat Neanderthalers konden praten. Kun je daar iets meer over vertellen?’

6) ‘Hoe weet jij dit eigenlijk allemaal? Op welke autoriteit baseer je je verhaal?’

- A op een artikel op een website met leuke weetjes, o.a. over Neanderthalers
- B op een wetenschapper die onderzoek heeft gedaan naar Neanderthalers
- C op een wetenschapsjournalist die allerlei onderzoek naar Neanderthalers heeft gelezen

Antwoorden

- 1) A
- 2) C
- 3) Vanuit een gevonden tongbeen konden wetenschappers afleiden dat ze een spraakorgaan hadden.
- 4) Er is geen specifiek taalgen, maar enkele genen helpen om klank te maken.
- 5) Ze konden vuur beheersen / controleren, en daarvoor moesten ze goed samenwerken. (En voor een goede samenwerking heb je vaak taal nodig.) Let wel: 'vuur maken' is fout. Om vuur te maken, hoef je immers niet te kunnen samenwerken.
- 6) C

BIJLAGE 4

Schattingen voor stellingen afzonderlijk
(hoofdstuk 6)

Tabel Parameterschatting van de gemiddelde score per stelling per conditie op de post-then-pre-test (S^2_k ; variantie tussen leerlingen; $S^2_{k_i}$; variantie tussen klassen; S^2_r ; residuele variantie)

Stelling	Docent-gestuurd		Zelfstandig			S^2_p ; S^2_k ; S^2_r
	gem. vóór	gem. na	gem. vóór	gem. na		
<i>Terugkijkgedrag</i>						
1. lees ik enkele woorden of woordgroepen terug.	2.96	3.64	3.02	3.53	a c	(.57; .00; .28)
2. lees ik hele zinnen terug.	2.84	3.58	2.91	3.39	a c	(.51; .01; .28)
3. lees ik een of meerdere alinea's terug.	2.23	3.22	2.35	2.99	a c	(.66; .02; .30)
4. stop ik bij een moeilijk woord dat ik niet ken.	3.21	3.69	3.18	3.48	a c	(.94; .00; .25)
5. lees ik veel elementen in de tekst opnieuw.	2.46	3.15	2.41	3.00	a	(.63; .01; .28)
<i>Inzet andere strategieën</i>						
6. scan ik vóór of ná het lezen de tekst.	2.73	3.50	2.55	3.25	a	(1.05; .06; .40)
7. let ik tijdens lezen op de hoofdzaken in de tekst.	2.50	3.40	2.66	3.25	a c	(.88; .04; .35)
8. pauzeer ik tijdens lezen om even na te denken.	2.20	2.78	2.21	2.69	a	(.85; .02; .30)
9. maak ik tijdens of na lezen aantekeningen, bijv. in de kantlijn.	1.30	1.73	1.35	1.66	a c	(.49; .00; .24)
10. formuleer ik hetgeen ik lees, in mijn hoofd om in eigen woorden.	2.45	3.18	2.44	3.07	a	(.81; .03; .32)
<i>Opbouwen begrip</i>						
11. lees ik de tekst zo dat ik hem naderhand in mijn hoofd kan samenvatten.	2.58	3.41	2.56	3.24	a c	(.76; .03; .38)
12. breng ik in mezelf de belangrijkste denkstappen van de tekst in kaart.	2.01	2.67	2.02	2.55	a c	(.69; .01; .28)
13. kan ik na lezen zelfstandig de hoofdgedachte ervan formuleren.	2.88	3.59	3.01	3.37	a c	(.66; .01; .28)
14. controleer ik na lezen af of ik de tekst goed genoeg begrepen heb.	2.94	3.84	2.97	3.65	a c	(.65; .01; .37)
15. denk ik dat ik na lezen de tekst voldoende begrijp.	3.25	3.85	3.33	3.67	a c	(.53; .01; .31)
<i>Dimensies van diep begrip</i>						
16. denk ik erover na of en hoe de tekst met bronnen is onderbouwd.	1.65	2.41	1.97	2.56	a c	(.64; .01; .33)
17. denk ik erover na of de schrijver ervan (of een bron) gezaghebbend is.	1.62	2.28	1.88	2.44	a	(.81; .02; .33)
18. denk ik erover na of en hoe ik een tekst voor een opdracht kan gebruiken.	2.61	3.27	2.55	3.12	a	(.86; .03; .29)
19. denk ik erover na of ik vanuit de gedachtegang zelf vervolgdienstapen kan zetten.	2.10	2.74	2.07	2.58	a	(.71; .01; .29)
20. denk ik na over wat de tekst me wil zeggen.	2.72	3.46	2.89	3.48	a c	(.84; .02; .33)

a: significant hoofdeffect voor/na de lessenserie; b: significant hoofdeffect conditie; c: significant interactie-effect voor/na lessenserie met conditie

BIJLAGE 5

Fit en passingsvergelijking van de modellen (hoofdstuk 6)

Tabel Fit en passingsvergelijking van tien modellen voor de score op de begripstoetsen (-2LL: -2log likelihood)

Model	-2LL	Modelvergelijking			
		Modellen	$\Delta\chi^2$	Δdf	p
1 intercept	13193.44				
2 + tijdstip van meting	11534.99	2 vs 1	1658.45	3	< .001
3 + interventie	11534.95	3 vs 2	.04	1	.84
4 + tijdstip interventie	11532.44	4 vs 3	2.51	1	.11
5 + conditie	11532.43	5 vs 4	.01	1	.92
6 + interventie * conditie	11528.46	6 vs 5	3.97	1	<.05
7 + percentage gecontroleerde antwoorden * interventie	10589.05	7 vs 6	939.41	1	<.001
8 + percentage gecontroleerde antwoorden * interventie * conditie	10588.52	8 vs 7	.53	1	.47
9 + niet terugkijken * interventie	10588.33	9 vs 8	.19	1	.66
10 + niet terugkijken * interventie * conditie	10587.82	10 vs 9	.51	1	.48

In bovenstaande Tabel is de fit en passingsverbetering weergegeven van tien modellen voor de score op de begripstoetsen. Als model 1 geldt een intercept: een gemiddelde score op de vier toetsen tezamen, met een schatting van de klassenvariantie en een variantie op leerlingniveau en error. Omdat de toetsen in maximumscore afwijken en op verschillende momenten zijn afgenomen, voegen we in model 2 het tijdstip van meting toe. Dit levert een passingsverbetering op – logischerwijs, gezien de wisselende maximumscore bij de vier toetsen (Model 2 vs Model 1: $\Delta\chi^2(3) = 1685.45$; $p \leq .001$). Zie de Tabel.

Interventie: Vervolgens verdisconteren we met model 3 het algemene effect van de interventie op de scores van de vier toetsen, wat niet tot een betere passing leidt (Model 3 vs Model 2: $\Delta\chi^2(1) = .04$; $p = .84$). Daarna verdisconteren we hierin het tijdstip van de interventie: in hoeverre is er een verschil in score waar te nemen tussen de eerste groep die de interventie ontving en de tweede groep die de interventie ontving binnen het *switching-replications*-design (model 4)? Dit leidt evenmin tot een betere passing (Model 4 vs Model 3: $\Delta\chi^2(1) = 2.51$; $p = .11$). Voor deze effectmeting van de scores binnen het *switching-replications*-design maken we hier gebruik van het *fixed-occasion*-model (Goldstein, 1987; Quené & Van den Bergh, 2004).

Conditie: Met model 5 bekijken we in hoeverre er een algemeen verschil in score aantoonbaar is tussen condities – ongeacht of leerlingen de interventie hebben gevolgd: hier

vinden we geen passingsverbetering (Model 5 vs Model 4: $\Delta\chi^2(1) = .01$; $p = .92$). Hierna bekijken we met model 6 in hoeverre dit verschil in score per conditie afhankelijk is van het feit of leerlingen de interventie hebben gevolgd: dit blijkt wel tot verbetering te leiden (Model 6 vs Model 5: $\Delta\chi^2(1) = 3.97$; $p < .045$).

Wijze van uitvoering: Ten slotte testen we in hoeverre de wijze van uitvoering door leerlingen van de experimentele lessenserie van invloed is op de scores. Daarvoor stonden ons twee controlevariabelen ter beschikking: het percentage gecontroleerde goede antwoorden in de experimentele lessenserie en het 'niet-terugkijken' naar de bestudeerde tekst. Toevoeging van een model waarin het percentage gecontroleerd goede antwoorden wordt verdisconteerd, leidt tot een zeer duidelijke passingsverbetering (Model 7 vs Model 6: $\Delta\chi^2(1) = 939.41$; $p < .001$); een interactie-effect met conditie leidt echter niet tot aantoonbare verbetering (Model 8 vs Model 7: $\Delta\chi^2(1) = .53$; $p = .47$). Toevoeging van twee modellen waarin het al dan niet terugkijken wordt meegenomen (model 9 en 10), heeft geen verbetering tot gevolg, niet als interactie-effect met interventie (Model 9 vs Model 8: $\Delta\chi^2(1) = .19$; $p = .66$), noch als interactie-effect met interventie en conditie (Model 10 vs Model 9: $\Delta\chi^2(1) = .51$; $p = .48$).

We rapporteren derhalve vanuit model 7 en laten model 8 tot en met 10 buiten beschouwing.

BIJLAGE 6

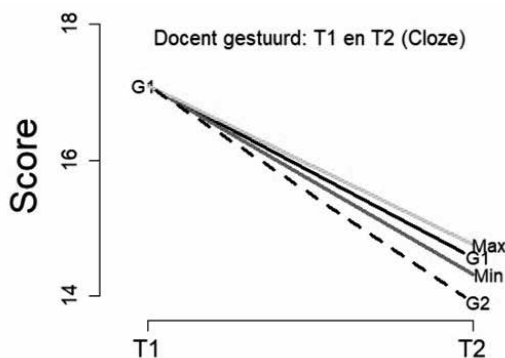
Effect interventie op tekstbegrip grafisch weergegeven (hoofdstuk 6)

Scores op de vier toetsmomenten grafisch weergegeven

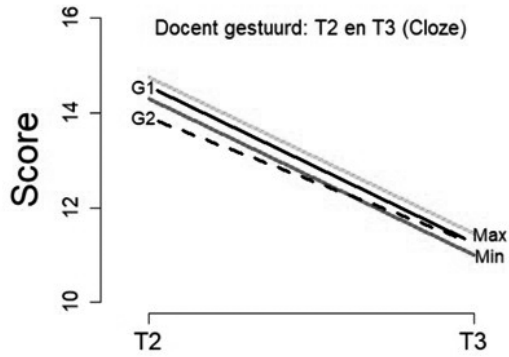
Opmerking vooraf. Wanneer de scores op de vier meetmomenten in één figuur worden gevisualiseerd, uitgesplitst voor conditie, voor moment van implementatie én voor wijze van uitvoering door de leerling, levert dat een overladen, nauwelijks leesbare figuur op.

Daarom zijn in deze bijlage drie figuren te vinden waarin de gemiddelde scores op de vier meetmomenten zijn opgesplitst in drie figuren, namelijk a) een voor het eerste en tweede meetmoment (figuur 1, voor de voor- en tussentoets), b) een voor het tweede en derde meetmoment (figuur 2, voor de tussentoets en cloze-natoets) en c) een voor het derde en vierde meetmoment (figuur 3, voor beide natoetsen).

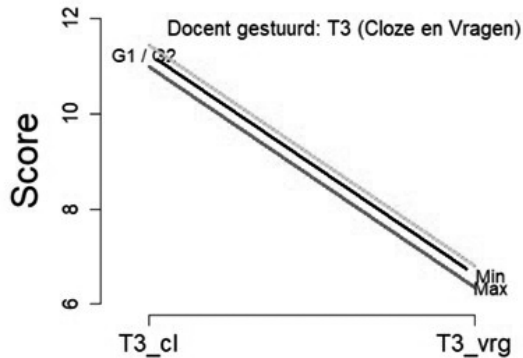
Daarbij is steeds ervoor gekozen alleen de conditie 'Docent-gestuurd leren' te veranschouwen, omdat hier sprake is van zowel een hoofdeffect van interventie, als een interactie-effect tussen interventie en de wijze van uitvoering door leerlingen; in de conditie 'Zelfstandig leren' ontbreekt dit hoofdeffect.



Figuur 1 Gemiddelde score voor de voor- en tussentoets, voor de conditie 'Docent-gestuurd leren', voor zowel de groep die de interventie na de voortoets ontving (G1) als de groep die de interventie pas na de tussentoets ontving (G2); voor G1 is daarbij de gemiddelde score veranschouwd van leerlingen die tijdens de lessenserie hun gegeven antwoorden minimaal controleerden ('Min') en maximaal controleerden ('Max').



Figuur 2 Gemiddelde score voor de tussentoets en cloze-natoets, voor de conditie 'Docent-gestuurd leren', voor zowel de groep die de interventie na de voortoets ontving (G1) als de groep die de interventie na de tussentoets ontving (G2); daarbij is de gemiddelde score veranschouwd van leerlingen die tijdens de lessenserie hun gegeven antwoorden minimaal controleerden ('Min') en maximaal controleerden ('Max').



Figuur 3 Gemiddelde score voor de cloze-natoets en vragen-natoets, voor de conditie 'Docent-gestuurd leren', voor beide groepen (G1 en G2); daarbij is de gemiddelde score veranschouwd van leerlingen die tijdens de lessenserie hun gegeven antwoorden minimaal controleerden ('Min') en maximaal controleerden ('Max').

SUMMARY

Focusing on deep reading

*Research on reading comprehension
of Dutch pre-university students*

1 Introduction

This study focuses on the reading comprehension skills of pupils at pre-university level, the better-skilled readers in Dutch secondary education (see Expertgroep Doorlopende Leerlijnen Taal, 2008, 2009; Gubbels et al., 2019; Van Silfhout, 2014; etc.). It does not, however, look at lower order reading skills, such as word decoding, but at higher order skills instead: deep reading, comprehension on the situation model level (Kintsch, 1998). We have studied this deep reading in tasks based on expository texts that are prevalent in course books for the Dutch language at pre-university level education in the Netherlands.

What prompted us to do this research? Firstly, content and teaching methods of reading comprehension in current course books for Dutch are, in many respects, based on understandings from the 1970s and 1980s. Secondly, the existing research into the skilled reader, i.e. pupils at pre-university level, in Dutch secondary education is relatively limited. Thirdly, when compared internationally, Dutch pupils seem less apt at evaluating and reflecting on texts. Deep reading skills should, therefore, get more attention in reading comprehension lessons (e.g. Van den Broek et al., 2021). Additionally, the school subject Dutch is currently being reformed and reading comprehension is a prominent focal point in this reform (Rijksoverheid, 2022). Hence, there seem to be enough reasons to study the reading comprehension skills of skilled pupils within the current ‘standard’ context of the school subject Dutch.

The purpose of our research is two-fold: on the one hand, we aim at a more precise description of what reading comprehension for pre-university pupils in the upper grades entails by analysing their behaviour in pre-reading a text and answering questions about that text; on the other hand, we aim to develop and test an effective alternative to the existing reading comprehension lessons for pre-university pupils. Our study is based on three research questions:

- a) How do pre-university pupils construct reading comprehension and how do they develop in this respect if we take standard reading comprehension lessons for the school subject Dutch as a starting point?
- b) To what extent is a text-with-questions an effective task to train reading comprehension skills for the school subject Dutch?
- c) What is an effective, process-oriented content and teaching method for reading comprehension lessons for the school subject Dutch?

Seven principles lie at the heart of this study:

- I. We want to focus on **deep reading**, on comprehension on the situation model level (Kintsch, 1998). In the second part of this book, where we discuss the design, we use Wolf's (2018) definition of deep reading: "the array of sophisticated processes that propel comprehension and that include inferential and deductive reasoning, analogical skills, critical analysis, reflection, and insight" (Wolf & Barzillai, 2009, p. 32).
- II. Our study group consists of pupils at pre-university level, the **skilled readers**, in Dutch secondary education (Expertgroep Doorlopende Leerlijnen Taal, 2008, 2009; Gubbels et al., 2019, Van Silfhout, 2014).
- III. We aim at making the indirect, rather elusive notion of text comprehension more tangible by **visualising** it: we use both eye-movement research as a descriptive research instrument and eye-movement videos as a didactic instrument (by using EMME's; Eye-Movements Modeling Examples). This will enable us to demonstrate to researchers, readers themselves, and pupils what text comprehension entails.
- IV. **Reading processes and question-answering processes** should be distinguished from one another. Within the school subject Dutch, statements are made on pupils' reading comprehension skills often based on their answers to comprehension questions; their thinking processes are mostly disregarded (e.g. Linthorst & De Glopper, 2015). In this book, we examine both the reading process and the question-answering process, and we examine the relation between those processes, as well as the relation to the reading outcome that pupils produce (i.e. their answers to comprehension questions).
- V. We will employ a **goal-directed** approach as to strategic reading behaviour: to what extent do readers succeed in attaining a previously set reading goal with the means they dispose of? The use of strategies is considered strongly dependable on the individual, the context and the goal. As such, we will employ a flexible interpretation of strategy instruction (Aarnoutse & Verhoeven, 2002).
- VI. Keeping all of the previous principles in mind, we strive for **ecological validity** in the following ways: a) embedding experiments as much as possible in regular assessments or lessons in Dutch secondary schools, b) involving as many teachers of Dutch as possible in our research project, c) fitting as much as possible into the existing practice of the school subject Dutch.
- VII. A more implicit principle is that we aim for an **optimal generalisability** of our research findings.

Since this study is two-fold, this book also consists of two parts. The first part, namely chapters 2 through 4, is descriptive and partly evaluative. We mainly used a *mixed method*-approach for this part. In chapters 2 and 3, we mostly used eye-movement research to describe and analyse pre-university pupils' pre-reading and question-answering behaviour when dealing with four texts with comprehension questions. In doing this, we pinpointed problem areas within the current teaching methods for reading comprehension in the school subject Dutch. Chap-

ter 4 addresses these problem areas in more detail. The second part of this book, namely chapters 5 and 6, concerns design and intervention research, reflecting the principles of observational learning (Bandura & Walters, 1977; Bandura, 1987) and recent theoretical understandings of text comprehension (e.g. Britt et al., 2018). On these premises, we suggest alternative design principles for reading comprehension lessons; we design and validate an alternative approach for lessons in reading comprehension for upper-level pupils primarily.

2 Answering the central research questions

In the following pages, we will discuss the three central research questions based on our findings in chapters 2 through 6:

2.1 Research question one

Our first research question was:

- a) How do pre-university pupils construct reading comprehension and how do they develop in this respect if we take standard reading comprehension lessons for the school subject Dutch as a starting point?

The results from chapters 2 and 3 generally support the legal descriptions in *Referentiekader Taal* (Framework of Reference Dutch Language Skills) of the current skill levels that pre-university pupils should reach. This Framework of Reference, incidentally, has been re-evaluated recently (Van den Broek et al., 2022). It presumes that pupils from level 2F (approximately the level in grade eight) onwards know the difference between major and minor points (Expertgroep Doorlopende Leerlijnen Taal, 2009), and chapter 2 confirms this presumption: in a standard text-with-questions task, tenth graders can distinguish between topic sentences and non-topic sentences. Chapter 3 also confirms this presumption.

In addition, chapter 3 shows cross-sectional differences between the lower and higher grades, which suggest a clear development of pupils' reading comprehension skills, in accordance with the description found in *Referentiekader Taal*. This chapter shows that pupils in tenth and twelfth grade read significantly faster than pupils in eighth grade and they re-read parts of a text more often: they also seem to read slightly more strategically than eighth graders. To our surprise, however, we were unable to demonstrate these differences between pupils in tenth grade and twelfth grade. These findings seem to be in line with the description in *Referentiekader Taal*, in which the step from basic level 2F to the more advanced level of 3F (approximately the level in tenth grade) seems larger than the step from level 3F to 4F (twelfth grade). The distinctions as outlined in *Referentiekader Taal* for the steps between the levels 2F and 3F (Van den Broek et al., 2022) are missing for the steps between the levels 3F and 4F. Moreover, the examination syllabi that describe the examination content and requirements for 3F (*havo*: upper pre-vocational level) and 4F (*vwo*: pre-university level) make little distinction between the two levels

(CvTE, 2022a, 2022b).

Furthermore, chapter 3 illustrates that there exists little distinction in the way pupils use pre-reading strategies. In grades eight, ten, and twelve alike, pupils generally read in a rather linear way, do not often re-read sentences and do not seem to make active use of (pre-)reading strategies. One of the core objectives for the school subject Dutch specifies that pre-university pupils, after their third year in secondary education, should be able to use and assess reading strategies (SLO, 2016a; Expertgroep Doorlopende Leerlijnen Taal, 2008). Yet, the pre-university pupils in our sample generally seem to employ these strategies only every so often when pre-reading a text, irrespective of their grade and a text's complexity that varied between levels 2F to 4F. This does not imply that pupils at pre-university level do not know how to use these strategies. On the contrary, it seems that the reading comprehension assignments in our experiments did not invite them to such strategic reading behaviour.

Chapter 3 also demonstrates that the pupils from the three grades that were part of this study do, however, show clear distinctions in their question-answering process. In comparison with the pre-reading process, the question-answering process is mostly (rather) thorough and consists of many steps. The distinction between eighth graders and twelfth graders is particularly striking. Firstly, twelfth graders spend less time reading the question than eighth graders, probably because they know how to interpret the question better. Secondly, after reading the question twelfth graders more often have an inkling of the correct answer, probably because they have constructed a better situation model of the text. Thirdly, twelfth graders know better where to find those parts of the text that are relevant to answer the question. They dedicate less time to parts that are irrelevant, probably because they have built a better text base of the text. Last but not least, twelfth graders more often answer the question correctly because they have built a better mental representation of the text and, subsequently, go through a more refined and appropriate question-answering process. The way in which tenth graders approach this task has more resemblance with twelfth graders than with eighth graders, which confirms once more that the step from 2F (grade eight) to 3F (grade ten) is larger than the one from 3F to 4F (grade twelve).

Selection is definitely an explanation for these differences between grades, as lower-level grade placement and grade retention lead to a more homogenous population in grade twelve compared to grade eight and, to a lesser extent, grade ten. In many respects, this cross-sectional description of reading and answer-giving behaviour suggests, nevertheless, a clear development of pre-university pupils' reading comprehension skills throughout all grades, which corresponds roughly with the description in *Referentiekader Taal*. Yet, this development is mainly apparent in the question-answering process and not in the pre-reading process. It is in that latter process that we notice that the rather linear, seemingly superficial, pre-reading process throughout all three grades leads to distinct text bases and situation models.

2.2 *Research question two*

Our second research question was:

- b) To what extent is a text-with-questions an effective task to train reading comprehension skills for the school subject Dutch?

The standard reading assignment ‘text with questions’ was the main focus of chapters 2 and 3. Based on eye-track-data, we were able to demonstrate, at the most, a weak correlation between pre-reading of a text on the one hand, and answering the question and the final answer to that question on the other hand. In chapter 2, we demonstrated that pupils in grade ten do make a distinction between topic sentences and non-topic sentences during pre-reading, but we were unable to demonstrate a correlation with their answer-giving behaviour and their ultimate answers. These findings are replicated to a large extent in chapter 3: again, we were able to demonstrate a weak correlation between reading of (non-) topic sentences and answer-giving behaviour and final answers.

Pupils seem to employ pre-reading strategies because they are told to do so and because in answering the questions, they better understand the question and know where to find the answer in the text. However, in chapter 3 we concluded from the think-aloud data that after reading the question, pupils in grade ten and especially in grade twelve knew better where to find the answer to the question in the text than pupils in grade eight. Still, only rarely did pupils, by their own account, have an (idea of an) answer after reading the question. Pupils who spent more time on the important sentences or less time on secondary issues showed hardly, or not at all, more appropriate answer-giving behaviour and did not answer the questions correctly more often compared to pupils who spent less time on the important sentences or more time on secondary issues.

Moreover, in chapter 3 we found traces of high task dependency and task familiarity: a small group of pupils (voluntarily!) indicated that they relied completely on text conventions (“The first question will undoubtedly be about paragraph 1”, “The first question was about paragraph 1, so the second one will no doubt be about paragraph 2 or 3”, etc.) As such, the results from chapters 2 and 3 seem to support the prevalent criticism on this task that pupils first and foremost learn tricks and learn to answer questions in a smart way (e.g. Hooegeveen & Bonset, 1998; Westhoff, 2012).

Chapter 4 confirms the conclusion that the answering phase in a reading comprehension task is crucial, as it is about consulting the text well when answering the questions and not about pre-reading a text intensively. In this study, we manipulated the pre-reading intensity in two separate texts with questions, in two experiments. In the first experiment (‘Consultability’), we wondered how being able to consult a text during the answering phase impacted the score for comprehension questions. That is to say, after pre-reading a text, to what extent do pupils who can consult the text attain a higher score than pupils who can no longer consult the text? In the second experiment (‘Intensity’), we wondered how the level of intensity of pre-reading the text impacted the score for comprehension questions. More specif-

ically, to what extent do pupils who have intensively studied the text attain a higher score than pupils who have either not studied the text or less intensively? In other words, these experiments were meant to find out whether studying a text beforehand, or consulting a text while answering the questions determines a pupil's score. To achieve that goal, we solely used questions at the situation model level, for two texts in both experiments. These texts had been used in the nationwide final exams Dutch for *havo/vwo* (upper pre-vocational and pre-university level).

In the first experiment ('Consultability'), we noticed that pupils who could not consult the text while answering the questions attained a considerably lower score than pupils who could. Thus, being able to consult the text has a large impact on the answer scores; something we anticipated.

In the second experiment ('Intensity'), we were unable to demonstrate a distinction in answer scores between the three groups that studied the text at varying levels of intensity; something we had not anticipated. So it is most likely that pre-reading a text intensively for this kind of assignment does not, or barely, lead to a higher answer score: being able to consult the text determines the answer scores. It seems obvious then that skilled readers such as pupils at pre-university level will only consult a text superficially.

Recent theoretical findings (notably the RESOLV model: Britt et al., 2018) also contributed to explain the unexpected results from this second experiment. Readers pre-read a text superficially to adequately interpret the questions and easily find the relevant parts to answer those questions. They will, therefore, ignore the instruction to *intensively* study the text because it is not until they read the questions that they see what the actual goal of the assignment is. Their activities in the answer phase determine whether pupils answer the questions correctly, that is to say, provided they can still consult the text. If they can no longer consult the text, which was the case in the first experiment in chapter 4, then readers will intensively study the text so as to answer the relatively complex comprehension questions correctly, we argued based on the model from Britt et al. (2018).

In chapter 4 we concluded therefore, that if we want to teach pupils how to read a text independently, we should employ this task with care. This task mostly teaches pupils to correctly answer questions about a text; they will most likely not learn how to read a text independently. Based on our research in chapters 2, 3, and 4, the focus on this task in reading comprehension lessons does not seem to be an obvious approach.

While this conclusion, strictly speaking, only applies to pupils at pre-university level, we can assume that it also applies to reading comprehension lessons in general. If skilful readers, such as pupils at pre-university level, adopt a highly pragmatic approach to reading a text and answering questions, there seems no reason to believe that less skilful readers would take a different approach. Since these groups of pupils are generally less motivated, concentrated and competent, this reasoning is likely to be even more valid for other levels.

2.3 Research question three

Our last research question was:

- c) What is an effective, process-oriented content and teaching method for reading comprehension lessons for the school subject Dutch?

In chapters 5 and 6 we described the principles for an alternative content and teaching method to be used in reading comprehension lessons for pupils at pre-university level. At the basis of these chapters lies the concept of deep reading (Wolf, 2018): reading a text attentively and repeatedly to facilitate the process of reasoning about, analysing and reflecting on its content. Other principles that inspired us were the principle of observational learning (e.g. Bandura, 1986; Braaksma, 2002), a more flexible interpretation of strategy-instruction (e.g. NAEP Reading Framework, 2017), recent understandings about the functional embedding of text comprehension (e.g. McCrudden et al., 2011), and the importance of being able to determine the level of usability and reliability of information (e.g. Kirschner, 2017), which has mainly been used in theories focusing on *multiple document comprehension* (e.g. Braasch et al., 2018). In addition, we embraced recent insights that skilful readers distinguish themselves from less skilful readers not only in their text comprehension, but also in their task model, planning, and approach of a reading comprehension task (Britt et al., 2018).

With all of these principles in mind, we developed a lesson series that in many respects can be considered a counterpart to the more common ‘text-with-questions’ approach. In many studies dedicated to observational learning this type of learning is generally alternated with learning by doing (e.g. Elving-Heida, 2019). That is why in this lesson series we developed lessons for learning by observing, in which pupils thought about (the construction of) text comprehension, and alternated these with practical lessons in which they applied what they had learned previously. In the lessons learning by observing, pupils described the reading process of their peers and of themselves, they reflected on the process and evaluated it. Thus, the effectivity of their reading process (and their use of reading strategies) automatically became a topic of discussion. On top of that, the texts were offered within a meaningful context, with a clear reading purpose. This way pupils could reflect on how the texts fulfilled their reading purpose. In other words, they reflected on the reliability and usability of texts. In the practical lessons, the reading tasks always served the purpose previously stated, their form changed to a large degree and they were to be carried out without being able to consult the text itself. Thus, the follow-up task became a tool to monitor the overall comprehension level, stressing the importance of studying a text independently prior to the task. In the last two lessons of the lesson series in chapter 6, pupils’ task models were discussed: how should a complex reading task be carried out to sufficiently complete a challenging reading task in school?

In chapters 5 and 6 we noticed that this alternative approach was, on the whole, effective. In both chapters, pupils indicated that after the lesson series they increasingly showed reading behaviour typical for deep reading. The participating teach-

ers and pupils were, generally speaking, also positive about the lesson series (“cool”, “an eye-opener”). Chapter 5 shows that, based on the test results, the pupils who had followed the lesson series had learned more than those in the control group who had followed the regular lessons. In chapter 6 we observed that the experimental lesson series had an effect on learning only if two conditions were fulfilled. First, pupils must conscientiously do all tasks in the lesson series. In both research conditions, the way pupils perform tasks determines how much they actually learn from the lesson series. Pupils that, for instance, do not check their answers, or only superficially, do hardly or not at all learn from the lesson series. Second, the teacher should hold class discussions during the lessons leaning by observing and should actively manage those discussions. Only in the research condition ‘Teacher-centred learning’ could we prove a general effect on the pupils’ learning abilities. If the teacher fails to monitor progress and lead reflective discussions in class, the individual engagement of pupils determines if and how much they learn from the lesson series.

Our conclusion is, therefore, that classroom activities such as this lesson series require an active role of the teacher and should be used with care otherwise.

3 Implications for science

In section 4, we discuss the implications of our study for teaching reading comprehension. First we look at the scientific implications of our research.

3.1 Methodological limitations

We discuss three methodological limitations. One: general research method. In chapters 2, 3, 5, and 6 we used a *mixed-method*-approach that combines quantitative with qualitative research (Creswell & Plano Clark, 2018). These *mixed-method*-studies are highly laborious, which significantly limited our research scope.

Two: the possibility of generalizing our findings. Experiments in chapters 2, 3, and 5 have all been conducted at a single secondary school only, which makes it problematic to draw general conclusions based on the findings. Replicating research findings in chapters does, however, largely overcome this problem.

Three: in our research groups, we did not distinguish between skilled and less skilled readers, even though that distinction offered interesting perspectives on its impact in other design research studies (e.g. Bouwer & Koster, 2015). Moreover, our study only encompassed short, non-fictional texts from Dutch magazines and newspapers, which are currently commonplace within the school subject Dutch.

3.2 Suggestions for future research

We examine four general suggestions for future research. One: characteristics of task-oriented reading. In chapter 4 we examined the task model of skilled read-

ers for a text-with-questions, customary for the school subject Dutch. We used the RESOLV model by Britt et al. (2018) as a starting point. The findings in chapter 4 enabled us to further define that model. Yet, there are ample grounds for future research into this particular reading assignment and other assignments as well.

Two: text-with-questions as means to assess reading comprehension test. Most theories on reading comprehension state that readers reach an adequate, deeper understanding of a text if they build a coherent situation model in keeping with that text (Kintsch, 1998; Magliano et al., 2018; Van den Broek, 2012; Van den Broek et al., 1999). In an ideal world, that is what an assessment of deep reading is able to measure (e.g. Magliano et al., 2018). Chapter 4, however, demonstrates that a regular test consisting of a text with questions does so inadequately. Thus, there seem to be sufficient reasons to look for alternatives for this type of test (see Sabatini et al., 2018, 2020).

Three: conditions for effective learning by observing. In line with Braaksma (2002), our studies in chapters 5 and 6 have accommodated the evaluative condition for learning by observing in the lesson series. Future research could focus on the extent in which these ‘reinforcing’ components are crucial for effective learning by observing.

Four: effectivity of types of models. Not only can EMME’s serve as an instrument of reflection for observational learning, as in chapters 5 and 6, but so can video recordings of peers that are thinking aloud or online recordings of reading or writing activities of models on a screen. An obvious future research project could examine which instrument would demonstrate the thinking process of models best, given the situation and learning domain.

4 Implications for the educational system

What does the study in this book mean for reading comprehension lessons in the school subject Dutch in secondary education? We discuss four implications.

Implication 1: Provide a meaningful context

Presently, teaching methods for reading comprehension in the school subject Dutch hardly account for the possible impact of a reading objective on readers’ understanding of a text. It seems, therefore, advisable to add a recognisable objective or a realistic context to reading comprehension assignments. This thesis argues to a) cautiously utilise the text-with-questions as a reading assignment, b) provide a meaningful context by integrating reading comprehension into other language skills or by linking it to the content, and c) vary the assignments and, in that variation, focus on deep comprehension.

Implication 2: Reflect on the comprehension process together

We recommend that teachers and pupils talk about the comprehension process more and about the relation that that process has with the end product. Thus, our recommendations are the following: a) focus more on the comprehension process of pupils themselves, b) turn reading into a dialectic activity, and c) reflect on texts using a set of functional analysis tools.

Implication 3: Compare, weigh and intergrate information between text(s) (parts)

In both scientific theories (Fox & Alexander, 2017) and international reading comprehension tests (OECD, 2019), comparing, evaluating, and synthesizing information from several sources about a specific topic has become widespread. It should, therefore, be more integrated in reading comprehension lessons for the school subject Dutch. This thesis states that teachers should a) devote more attention to multi-text comprehension, b) offer a wide variety of (digital) texts, and c) give more attention to evaluating and reflecting on texts in their lessons.

Implication 4: The teacher's role in the learning process is crucial

Chapter 6 shows that teachers play a crucial role in the pupils' learning process. Notably, reflective discussions and evaluation in the classroom, monitored by teachers, can be a valuable instrument in learning by observing. We wonder to what extent teachers are currently prepared for that role. That is why we recommend to train teachers in process-oriented classroom reflection.

DANKWOORD

“Abstracties: mazen waar de werkelijkheid doorheen zakt”, schreef Maria Vasalis ooit. Ik kon dit boek pas schrijven toen ik de waarheid van dit zinnetje min of meer ging begrijpen. Elk denkbouwwerk heeft zijn fundamentele gebreken, en dat geldt ongetwijfeld ook voor dit boek. Die noodzakelijke onvolmaaktheid maakt het creëren van een helder geordend, stevig denkbouwwerk echter nog niet irrelevant. Hemel zij dank is dit boek geen tijdloos opus, geen *monumentum aere perennius*. Dan zou het ook nooit geschreven zijn.

Vooropgesteld: dit onderzoek zou er eenvoudigweg niet geweest zijn zonder Huub van den Bergh, die jaren terug erop aandrong dat ik onderzoek ging doen. Huubs humor, vertrouwen en steun, zijn relativiseringsvermogen en schalkse lichtheid, zijn kennis, scherpte en overzicht: telkens als het nodig was, stond hij voor me klaar. Ik heb onze samenwerking vijf jaar lang ononderbroken als een plezier ervaren. Huub, lang geleden omschreef je bij een andere promovendus van je het voleindigen van een promotieproject als het “behalen van het academisch middenstandsdiploma” – wat een plezier bij jóú dit diploma te hebben mogen halen.

Een halfjaar na het starten van m'n onderzoek sprak ik Gerdineke uitgebreid, en al vrij snel werd ze copromotor. Gerdineke, scherp, alert, grappig, spin in het web van het schoolvak Nederlands – altijd op de hoogte van wat waar speelt, vol onderbouwde ideeën wat er anders kan en moet: zonder haar zou de relevantie en diepgang van het onderzoek in dit boek zeker minder zijn geweest.

Aanzienlijk later sloot ook Erwin aan bij m'n project. Erwin en ik kennen elkaar allang: ik zat ooit kort als student bij hem in de collegebanken. Erwin, dank voor de warme, inzichtrijke gesprekken van het afgelopen anderhalf jaar waarin ik door jouw scherpzinnige, bezonnen vragen mezelf kristalhelder hoorde zeggen wat ik daarvoor écht nog niet ten volle begrepen had.

Twee andere mensen moeten hier ook genoemd worden. Allereerst Uriël Schuurs die bij de eerste drie studies in dit boek als begeleider optrad en enkele malen belangrijke aanvullingen en aanscherpingen gaf bij de onderzoeksopzet: vooral in de studie in hoofdstuk 4 heeft Uriël cruciale inbreng gehad. Daarnaast dank ik graag Iris Mulders die voor de studies in hoofdstuk 2 en 3 in allerlei programma's ontwierp die voor onze eye-tracking-analyses noodzakelijk waren.

Verder dank ik hier alle collega-docenten Nederlands die aan dit boek hebben meegewerkt. Geweldig dat ze tijd ervoor vrijmaakten en moeite ervoor deden in hun drukke schoolse bestaan! Enkelen noem ik graag bij naam: allereerst mijn vriend Twan Robben, die altijd te porren was voor onderzoek en meedenken. Ik

denk ook graag aan twee collega's van het Sint-Janslyceum, mijn oude vertrouwde school: Inge van Luijtelaar, met haar feilloze organisatievermogen, en Eva Bongers, die me zoveel nakijkwerk uit handen heeft genomen. Sowieso moet ik mijn oude school bedanken voor de ruimhartigheid waarmee ze mij de vrijheid gaf onderzoek te doen. Graag dank ik ook de collega's van de Pierson, met name Hilde Stinstra, voor hun herhaalde bereidheid om aan m'n onderzoeken deel te nemen. Ook Yke Schotanus, Janneke Stuulen en Coen van 't Veer wil ik hier graag bij name noemen, omdat ze op cruciale momenten een helpende hand toestaken.

Van de gesprekken die ik naar aanleiding van mijn onderzoek heb gehad, denk ik met erg veel plezier terug aan het lange onderhoud dat ik met Paul van den Broek had. Ook dank ik graag Gert Rijlaarsdam die me enkele malen zeer bruikbare suggesties gaf. En Corina Breukink noem ik ook graag, met wie ik enkele malen lange discussies had over hoe en waarom van onderzoek naar tekstbegrip van zakelijke én literaire teksten.

Dank ook aan Peter Coopmans en Bernadette de Zeeuw – samen het warm kloppend hart van het Dudoc-Alfa-programma. De weekenden in de Reehorst, samen met de andere promovendi (waarbij ik met name de andere neerlandici moet noemen!), waren onvergetelijk.

Ten slotte dank ik graag mijn huidige werkgever, Cito, die ertoe bereid bleek dit boek uit te geven en die me de ruimte bood dit project tot een goed einde te brengen. Inge Hermsen, wat leuk dat je hier nog aan hebt meegewerkt in de afrondende fase: dank voor het doorvlooiën van het manuscript. Suzanne van den Berg bedank ik ook erg graag: geweldig dat je de vertaling van de samenvatting op je wilde nemen. En Roelien Linthorst en later Wilma Vrijs: dank dat jullie geduld met me hadden en zo goed met me mee hebben gedacht.

En dán: mam, hoe zou ik ooit dit hebben kunnen maken als jij niet vroeger eindelijk de tijd zou hebben genomen me duidelijk te maken dat leren goed voor me is? Jammer dat ons pap niet mee heeft kunnen maken dat ik dit grote project afmaakte.

En ten slotte het kernteam thuis. Rein en Clara: sorry dat papa soms 's avonds zo wazig voor zich uitstaarde na een dag statistische berekeningen en in onderzoeksferen verwijlend de spruiten schildde. Wat fijn dat jullie altijd gewoon weer onbekommerd om me heen snaterden en speelden! En ja, als laatste noem ik haar dan: Channa, mijn lief, die me altijd weer met beide benen op de grond zet en zonder wie alles (ja, alles) zo veel minder leuk zou zijn.

NAMENREGISTER

Register op auteursnaam, voor de pagina's 1-201.

- Aalsvoort, M. van der 17
Aarnoutse, C. 27
Acket, J.M. 13
Afflerbach, P. 17, 27, 28, 34, 48, 56-58, 83,
133, 134, 139
Agarwal, P.K. 92
Alderson, J.C. 181
Alexander, P.A. 193
Allen, L.K. 129, 132
Almasi, J.F. 11, 18, 117
Alvarado, J.M. 97
Andersson, R. 21, 26, 34, 37, 39-40, 50, 57,
62, 64, 81-82, 114, 173
Andre, T. 56
Anmarkrud, Ø. 84, 110
Ardoin, S.P. 91
Artelt, C. 108
Ballering, C. 21
Bandura, A. 29, 113, 119, 134, 135, 157,
166, 171, 183, 190
Baron, N.S. 25, 129, 132
Barzillai, M. 25, 129, 132
Bax, S. 26, 34, 50, 57, 58, 60, 81-83, 91,
106, 114, 173, 174, 177, 181
Bergh, H. van den 131, 174
Berry, S.L. 183
Betz, D.L. 120, 145
Biggs, J. 176
Bimmel, P. 134
Blauw, A. de 192
Bogaerds-Hazenbergh, S. 27, 28, 90, 130,
140, 158-159, 192, 197, 198
Bogaert, N. 55
Bolt, J. 11
Bonset, H. 17, 20, 23, 25-27, 35, 50, 55, 56,
83, 87, 130, 158, 169, 186-188, 193, 196
Bosch, J.H. van den 13
Bouhnik, D. 136, 156, 158
Bouwer, R. 4, 113, 157, 175, 182, 184, 190
Braaksma, M. 23, 26, 27, 114, 115, 119,
134, 139, 142, 182-184, 198
Braasch, J.L. 131, 157, 171, 187
Braet, A. 16, 19, 200
Branden, K. van den 21, 83, 87, 192, 193
Bråten, I. 110
Breij, B. 200
Breukink, C. 48, 80, 107, 116, 125, 184
Briesch, A. 144
Brillenburch Wurth, K. 14
Brinke, S. ten 16
Britt, M.A. *passim*
Broek, A. van den 167
Broek, P. van den 20, 21, 88, 89, 118, 130-
134, 141, 180-181, 187, 193, 201
Brown, A.L. 134
Brown, R. 134
Bruin, E. de 35
Brunfaut, T. 26, 40, 50, 58, 60, 81, 82, 83,
91, 106, 174, 177, 181
Cain, K. 91, 146, 181
Canton, J. 28, 34, 56, 90, 96, 169, 197
Carroll, C. 175
Cataldo, M.G. 90, 91
Catrysse, L. 129
Cerdán, R. 33, 34, 38, 58, 59, 81, 82, 83, 87,
90, 91, 180
Cervetti, G. 26
Clowes, R.W. 133
Cohen, A.D. 26, 91, 153, 181
Copier, J. 19
Cort, F. de 30
Couzijn, M.J. 23, 26, 27, 84, 109, 113, 114,
116, 123, 134, 135, 157, 159, 182, 184
Cozijn, R. 34, 39, 57, 63, 172, 173
Creswell, J.W. 143, 174, 175
Curriculum.nu 21, 132, 159
Cutting, L.E. 21
CVEN 15, 17

- CvTE 19, 130, 167, 178, 199
 Damhuis, R. 192
 De Milliano, I. 21, 55, 87
 Dera, J. 21, 195
 DESI 129
 Diels, P.A. 16, 191
 Dijk, T.A. van 88, 132
 Dijk, Y. van 20, 200
 Dis, L.M. van 15
 Doeselaar, N. van 195
 Donker, A.S. 134
 Dönszelmann, S. 186
 Drop, W. 17, 18, 23, 27, 56, 59, 87, 131,
 Duchowski, A.T. 34, 57
 Duke, N.K. 34, 35, 48, 56, 58, 131, 134, 191
 Dunlosky, J. 91
 Durkin, D. 18, 130, 158, 188
 Edmonds, W.A. 173, 174
 Elbaz, E. 132
 Elbro, C. 146
 Elving-Heida, K. 113, 144, 157, 171, 175,
 182, 183, 184,
 Embretson, S.E. 81, 174,
 Ent, W. van den 15
 Epstein, W. 183
 Evers-Vermeul, J. 113, 182, 190, 193, 195
 Expertgroep Doorlopende Leerlijnen Taal
 11, 19, 22, 25, 33, 36, 47, 56, 58, 59, 80,
 130, 167, 178, 199
 Fang, Z. 130
 Ferrer, A. 88, 92, 94, 108, 139, 182
 Fisher, D. 18, 19, 84, 130, 140, 141, 192
 Fox, E. 193
 Frank, H. 11, 23, 35, 87, 130, 158, 161,
 178, 190
 Frey, N. 18, 19, 84, 130, 140, 141, 192, 199
 Fullerton, S.K. 11, 17, 117, 131, 133, 138,
 191
 Gelderen, A. van 87, 159
 Gellert, A.S. 146
 Gerjets, P. 34, 57
 Gil, L. 91, 109, 133, 140, 180, 194
 Glenberg, A.M. 183
 Glopper, K. de 26, 27, 55, 60, 87, 109, 113,
 130, 131, 190, 200, 201
 Goldman, S.R. 25, 136
 Graesser, A.C. 56, 82
 Graham, S. 132, 134
 Grenner, E. 182, 183, 184
 Grezel, J.E. 87
 Griffioen, J. 16, 18, 23, 185, 199
 Groenendijk, R. 23, 87, 107
 Groenendijk, T. 182
 Gubbels, J. 11, 20-22, 25, 55, 56, 83, 93,
 109, 116, 129, 131, 159, 178, 189, 193, 201
 Guthrie, J.T. 133
 Haas, E. 16
 Hafner, C. 133
 Halamish, V. 132
 Hamilton, R.J. 33, 34
 Hamm, D.N. 90, 146, 181
 Harris, K.R. 134
 Hattie, J. 134, 135, 156, 161, 197
 Hayles, N.K. 25, 129, 132, 137
 Hebert, M. 132
 Heemskerk, K. 187
 Helder, A. 89, 118, 133, 134, 141
 Hendrix, T. 18, 23, 90, 109, 176, 186
 Hermida, D. 25, 136
 Higgs, K. 92, 139, 182
 Hill, L.G. 120, 145
 Holland, N. 14
 Holmqvist, K. 21, 26, 34, 37, 39-40, 50, 57,
 62, 64, 81-82, 114, 173
 Hoogeveen, M. 18, 23, 50, 56, 83, 87, 107,
 130, 146, 158, 169, 187, 188
 Houtveen, A.A.M. 21, 124, 192, 193
 Houtven, T. van 25, 136
 Hulshof, H. 12, 13, 14, 15, 16, 18, 23, 90,
 109, 186
 Hyönä, J. 35, 47, 48, 50, 58, 173, 174
 Inspectie van het Onderwijs 21, 130
 Jackson, M.D. 57
 Jarodzka, H. 141, 184
 Johnston, P. 92, 108, 182, 190
 Jones, R. 133
 Jonz, J. 119, 146
 Kalmthout, T. van 12, 13, 14
 Kalshoven, F. 19
 Kamalski, J. 35, 51, 83, 108, 119, 146, 175,
 177
 Keehnen, T. 27, 84, 113, 114, 116, 123,
 157, 159, 182, 184
 Keenan, J.M. 146
 Kellogg, R.T. 113, 183
 Kennedy, T.D. 173, 174

- Kintsch, W. *passim*
 Kirschner, P.A. 131, 161, 193, 195
 Kitsantas, A. 190
 Klaver, M.J. 20
 Kleijn, S. 146
 Knoop-van Campen, C.A. 184
 Koster, M. 157, 175, 184
 Kraal, A. 57, 172, 173,
 Kramer, W. 14
 Kreeft, H. 23, 185,
 Kreft, I.G. 174
 Kuijpers, C. 57
 Kuperman, V. 173
 Lam, T.C. 120, 145
 Land, J. 21, 35, 38, 51, 55, 62, 83, 84, 96
 Langeveld, J. 15, 16, 191
 Lapp, D. 15, 116, 160
 Leest, J. 15
 Leeuw, L. de 57, 172
 Leeuw, B. van der 55, 82
 Liaw, S.S. 136
 Linderholm, T. 180, 187
 Linthorst, R. 26, 27, 55, 58, 60, 87, 109,
 113, 130, 131, 190
 Liu, Z. 129, 132
 Llorens, A. 114, 115, 118, 134, 135, 141,
 157, 184
 Loo, J. van der 160, 182, 184, 185
 Lorch, R.F. 34, 35, 58, 140, 173
 LTN [Levende Talen Nederlands] 97, 107
 Luijten, T. 23, 185
 Lysynchuk, L.M. 134
 Magliano, J.P. 34, 35, 57, 119, 132, 141,
 181, 186
 Maki, R.H. 86, 91, 183
 Mangen, A. 132
 Marcus, T. 136, 156, 158
 Marlen, T. van 26, 114, 116, 124, 141, 184
 Martínez, R.A. 136, 158
 McClelland, J.L. 57
 McCray, G. 26, 40, 50, 58, 60, 64, 81, 82,
 83, 90, 91, 106, 174, 177, 181
 McCrudden, M.T. 19, 22, 108, 109, 133,
 135, 140, 157, 171
 McNamara, D.S. 35, 38, 56, 57, 58, 62, 96,
 119, 129, 132, 140, 141, 186
 Meesterschapsteam Nederlands 132, 159
 Meestringa, T. 15, 34, 35, 55, 56
 Menninghaus, W. 200
 Merkx, P. 11, 35
 Merrill, M.D. 187
 Miller, B.W. 173
 Meulen, F. van der 177
 Ministerie van Onderwijs, Cultuur &
 Wetenschappen 130
 Moeken, N. 191
 Moormann, J.G.M. 14
 Mortel, K. van de 21
 Mross, E.F. 35, 50
 Mullis, I.V.S. 131, 195
 NAEP [National Assessment of Educatio-
 nal Progress] 134, 171
 Narvaez, D. 34
 Nicolaas, M. 20, 21, 129, 130, 131, 132
 Niemi, P. 35, 47, 50, 58, 173, 174
 NJI [Nederlands Jeugd Instituut] 30
 Norris, S.P. 83
 Nyström, M. 38, 62
 O'Reilly, T. 38, 62, 90, 91, 96, 108, 180
 Oakhill, J. 90, 91, 181
 Ockenburg, L. van 84
 OECD 19, 20, 108, 129, 131, 135, 157, 159,
 182, 195
 Okkinga, M. 27
 Oller, J. W. 119, 146
 Onderwijsraad 20, 21, 25, 83, 107, 116,
 125, 129, 130, 160, 161, 165, 189
 Onna, M. van 33
 Oostdam, R. 134
 Open Science Collaboration 177
 Ozuru, Y. 34, 82, 88, 91, 92, 139
 Palincsar, A.S. 134
 Paris, S.G. 140
 Parreren, C.F. van 17
 Payne, S.J. 90
 Pearson, P.D. 26, 34, 35, 58, 90, 131, 134,
 146, 153, 181, 182
 Pereira, C. 20, 21, 129, 130, 131
 Perin, D. 132
 Perfetti, C.A. 24, 25, 34, 35, 56, 57, 58, 118,
 136, 141, 194
 Piper, A. 25, 129, 132
 Plano Clark, V.L. 143, 174, 175
 Popham, W.J. 83, 188
 Pressley, M. 17, 27, 28, 34, 35, 48, 56, 57,
 58, 83, 133, 134

- Pronk-van Eunen, M. 11, 56, 130
 Quené, H. 174
 Raad voor Cultuur 20, 21, 25, 83, 107, 116,
 125, 129, 130, 159, 160, 165, 189
 Raad voor Nederlandse Taal en Letteren
 25, 136
 Raedts, M. 184
 Ravesloot, C. 15, 34, 35, 55, 56
 Rawson, K.A. 91
 Rayner, K. 39, 43, 63, 173
 Reader, W.R. 90
 Reber, R. 200
 Riemens-Feteris, E.S.F. 16
 Rigney, A. 14, 200
 Rijksoverheid 21, 130, 165
 Rijlaarsdam, G. 23, 26, 27, 113, 134, 182,
 184
 Robben, T. 25, 33, 55, 177, 189
 Robinson, F.P. 17
 Roediger III, H.L. 92
 Rogiers, A. 134
 Ros, B. 193
 Rosenblatt, L.M. 14
 Rouet, J.-F. *passim*
 Rupp, A.A. 26, 81, 91, 174, 181
 Sabatini, J. 108, 181, 182
 Salmerón, L. 59, 87, 90, 114, 115, 118, 134,
 135, 141, 156, 157, 175, 184
 Sanford, A.J. 81
 Saudargas, R. 144
 Scarborough, H.S. 146
 Schaffner, E. 92
 Schiefele, U. 92
 Schoot, M. van der 57, 173
 Schooten, E. van 159
 Schotter, E.R. 58
 Schram, D. 21, 55
 Schraw, G. 19, 133
 Schroeder, S. 88, 92, 94, 108, 117, 125,
 139, 182, 190
 Schunk, D.H. 114, 190
 Schuppert, L. 131
 Schuurs, U. 31, 53, 84, 85, 200
 Shadish, W. 137, 138, 175
 Sijtsma, K. 97
 Silfhout, G. van 11, 21, 22, 25, 55-57, 157,
 173, 174, 176, 178
 Slagter, P. 16
 SLO 15, 35, 50, 87, 107, 130-132, 135, 158-
 159, 165, 167, 192, 193
 Smeets, E. 135, 161, 196
 Son, E.H. 18, 192
 Sosnoski, J. 129, 132
 Stafura, J. 24, 57, 118, 141
 Steenbakkens, J. 200
 Steensel, R. van 4, 21, 124, 130
 Stellwag, H.W.F. 15
 Stemvers, F.G.A. 14
 Stevens, N. 48, 80, 107
 Stronks, E. 200
 Sturt, P. 81
 Traub, R.E. 91
 Trizano-Hermosilla, I. 97
 Truijens, A. 19, 20
 Udo, G.J. 136, 158
 Upton, T.A. 26, 91, 181
 Vakdidactiek Geesteswetenschappen 22
 Van Dyke, J.A. 173
 Ven, P.H. van de 12, 13, 16, 17
 Verhoeven, L. 27
 Vermeij, L. 177
 Vernooy, K. 189
 Vidal-Abarca, E. 90, 91
 Vos, B. de 11, 19, 56, 83, 130, 187
 Vries, J.H.L. de 17, 18, 27, 131
 Walczyk, J.J. 27
 Wallot, S. 57
 Walters, R.H. 135, 157, 190
 Warps, J. 25, 136, 160
 Welgraven, E. 19, 56, 83, 187
 Wesdorp, H. 56, 87
 Westhoff, G. 18, 23, 130, 158, 169
 Wetering, T. van de 23, 87, 107, 187
 Wetzel, C.D. 81, 174
 Wijekumar, K.K. 130
 WHO [World Health Organisation] 19,
 195
 Wilkerson, I.A.G. 18, 192
 Wimmer, H. 57
 Witte, T. 15
 Wolf, M. 25, 107, 116, 129, 132, 133, 134,
 135, 137, 157, 171, 189
 Yeari, M. 33, 34, 35, 47, 50, 56, 58, 80, 89,
 133, 173, 180, 181, 187
 Zimmerman, B.J. 190

AUTEURSINFORMATIE

Patrick Rooijackers (Deurne, 1978) studeerde in 2002 cum laude af aan de opleiding Nederlandse Taal en Cultuur van de Universiteit Utrecht, met als specialisatie moderne letterkunde. Na zijn afstuderen werkte hij een jaar lang aan de afdeling moderne letterkunde van de Rijksuniversiteit Leiden. In deze jaren publiceerde hij over Nederlandse literatuur in *Vooys*, *tijdschrift voor letteren*, *Liter* en *Tijdschrift voor Nederlandse taal- en letterkunde*.

In 2004-2005 volgde hij aan de universitaire lerarenopleiding van (IVLOS) de Universiteit Utrecht de opleiding tot eerstegraads docent Nederlands. Vanaf 2004 tot 2022 was hij werkzaam als docent Nederlands eerstegraads aan het Sint-Janslyceum in 's-Hertogenbosch, waar hij onder meer mentor was van vwo 5 en vwo 6 en lid was van de beleidsgroep van de vwo-afdeling.

Van 2005 tot 2009 was hij bestuurslid (en vanaf 2006 vice-voorzitter) van de vakvereniging van docenten Nederlands, *Levende Talen Nederlands*. In deze periode was hij ook redacteur voor het schoolvak Nederlands in *Levende Talen Magazine*, waarin hij publiceerde over didactiek en ontwikkelingen binnen het schoolvak Nederlands.

Sinds 2007 is hij betrokken bij de examinering rond het schoolvak Nederlands. Van 2007 tot 2016 was hij verantwoordelijk voor de vaststelling van de Centraal Schriftelijk Examens (CSE) Nederlands havo en vwo, als lid van de vaksectie Nederlands havo/vwo bij het College voor Toetsen en Examens (CvTE). Sinds mei 2021 is hij werkzaam als toetsdeskundige bij Stichting Cito. Vanuit Stichting Cito zit hij sinds september 2022 ook aan als adviseur toetsing bij de advieskring van de vakvernieuwingscommissie (vvc) voor het examenprogramma van het schoolvak Nederlands in de bovenbouw.

In april 2017 werd hem een onderzoeksbeurs toegekend vanuit het Dudoc-Alfa-programma, een programma dat eerstegraads docenten de mogelijkheid biedt een promotieonderzoek te doen. Vanuit deze beurs werd hij vanaf september 2017 als promovendus aangesteld aan de Universiteit Utrecht. Het onderzoek dat in dit boek gerapporteerd is, komt uit dit project voort.

Er leven anno 2023 veel zorgen over het dalend tekstbegrip van Nederlandse leerlingen. Volgens critici komt dit onder meer doordat docenten Nederlands “tekstbegrip vaak alleen nog maar trainen met vragenreeksen bij teksten”. Hiermee leren leerlingen inderdaad niet zelfstandig te lezen. Docenten kunnen beter realistische, doelgerichte leestaken gebruiken en klassikaal nadenken over de vraag hoe leerlingen leestaken aanpakken. Dit zijn de belangrijkste conclusies van *Oog voor diep begrip: onderzoek naar het tekstbegrip van vwo-leerlingen*.

In het eerste deel van dit proefschrift wordt het tekstbegrip van vwo-leerlingen geanalyseerd bij ‘teksten met vragen’. Tijdens oogbewegingsonderzoek blijken ze teksten vooraf oppervlakkig lezen; het vooraf-lezen en het beantwoorden van begripsvragen zijn grotendeels gescheiden denkprocessen. Voor een goed antwoord bij deze taak maakt het niet uit of vwo-leerlingen teksten vooraf intensief bestuderen, wél of ze bij de vragen de teksten kunnen raadplegen. Met deze taak lijken vwo-leerlingen dus vooral het beantwoorden van begripsvragen te oefenen, niet het zelfstandig bestuderen van teksten.

In het tweede deel van dit proefschrift staat een alternatieve lessenserie over ‘diep begrip’ centraal, gebaseerd op recent internationaal onderzoek. Het zelfstandig, realistisch, doelgericht bestuderen van teksten staat voorop; vwo-leerlingen reflecteren uitgebreid op de aanpak van lastige leestaken. De lessenserie blijkt effectief: mits docenten de reflectie leiden, laten vwo-leerlingen naderhand een hoger tekstbegrip zien en ze geven aan hun leesgedrag positief te hebben veranderd.

Het proefschrift sluit af met elf aanbevelingen om leeslessen Nederlands betekenisvoller, eigentijds en reflectiever te maken. Werk in de leesles procesmatiger, heb meer oog voor evaluatie en reflectie en gebruik daarbij een breed palet aan teksten en taken.